Transport of
hydrothermal
iron, manganese
and aluminium
across the
South Pacific

PAGES 160 & 200

HOT
METAL

The HIV epidemic can be stopped

Mounting evidence that rapid treatment with antiretroviral drugs dramatically reduces HIV transmission must be acted on fast if a target date for curbing the epidemic is to be met.

As scientists prepare to meet in Vancouver, Canada, for the annual meeting of the International AIDS Society (IAS) on 19–22 July, many argue that the end of the AIDS epidemic could be in sight. A mass of convincing data, they say, shows that the universal roll-out of antiretroviral treatment provides a means to stop HIV — but only if the world acts fast.

The optimism is due to the apparent success of the 'treatment as prevention' approach. Treating people with antiretroviral drugs as soon as possible after their diagnosis, it seems, not only prevents death and disability due to the disease but also prevents virus transmission. In 2014, the United Nations Joint Programme on AIDS (UNAIDS) drew on this concept to set the '90-90-90' goals, which envisage diagnosing and effectively treating 90% of people infected with HIV to eliminate the disease as a public-health threat by 2030.

In a report last month by a UNAIDS–*Lancet* commission, experts estimate that there is a five-year window of opportunity to make or break the 90-90-90 goals (see go.nature.com/ztqoj1). They note that the number of new infections is now declining year on year as more and more people receive antiretroviral treatment. As of 2013, nearly 13 million people were receiving antiretroviral drugs, a roughly tenfold increase over the previous decade. Should this trend continue, the Millennium Development Goal set in 2011, to get 15 million people on treatment by the end of 2015, will be exceeded.

Will this trend continue? There are 35 million people living with HIV, all of whom will eventually need antiviral therapy. Yet provision is too slow, the commission points out. It estimates that, if treatment is made accessible to new patients at the same rate as today, population growth in southern Africa will see the number of new infections and AIDS deaths per year creep up again by 2020. But if countries accelerate the provision of treatment in the next five years, the commission says, the goal of stopping the epidemic by 2030 is within reach.

Getting there will take a massive financial investment — as much as US$36 billion annually, compared with current investment of $19 billion per year. That represents as much as 2.1% of the gross domestic product of some affected nations.

Coaxing forth that level of investment in an age of austerity will be difficult. But, by modelling the economic gains of people remaining healthy and productive members of society, the commission estimates that countries with large HIV burdens will benefit from their increased spending.

As we report in a News Feature on page 146, scientists are also conducting research in 'implementation science', showing how to better provide treatment. And researchers are right to highlight results that definitively support an increase in investment as a means both to preserve health and to contain the epidemic by suppressing transmission.

The first major evidence supporting treatment as prevention came in 2011. A study called HPTN 052 found that providing treatment immediately on diagnosis to the infected partner of a couple regardless of whether his or her blood-cell count showed low numbers of the CD4 type of T cell — the usual marker of disease progression and indication for antiretroviral therapy — cut the risk that this person would transmit the virus to the uninfected partner by 96%.

Open questions, such as whether the approach would work in other settings, have now largely been answered. In February, the TEMPRANO trial in more than 5,000 people in Côte d'Ivoire reported that immediately starting antiretroviral treatment cut the risk of death and serious illnesses, such as tuberculosis and bacterial infections, by 44%. In May, the START trial, involving 4,685 people in 35 countries, was stopped early after reporting that immediate treatment cut the risk of serious illness or death by 53%. The trend was seen across low-, middle- and high-income countries.

> *"Treatment works for individual and public health, and for the public–health purse."*

On the basis of these and other results, the World Health Organization is considering revising its guidelines to recommend immediate provision of antiretroviral therapy to all people infected with HIV, not just to specific groups. The evidence for such a shift could be strengthened at the IAS meeting: the HPTN 052 trial will report whether the dramatic drop in transmission has held up in the longer term, and START will report its full results (the May results were preliminary).

Altogether, the evidence bolsters the case that the world now has the tools at hand to eliminate the HIV threat. As conference co-chair Julio Montaner of the University of British Columbia, Vancouver, argues: "Treatment works for individual and public health, and for the public-health purse. As a policymaker, you have nowhere to hide." ∎

A numbers game

Institutions must be plain about research metrics if academics are to engage with them.

Scientists like to grumble about the peer-review system for judging research quality, but there is one sure way to make most of them defend it: suggest that peer review should be replaced with numerical measures of academic output.

A major UK report on the use of such research metrics this week reinforces this defence of the status quo (see go.nature.com/smbaix). Metrics, it concludes, are not yet ready to replace peer review as the preferred way to judge research papers, proposals and individuals.

Even if such metrics do not replace peer review in all situations, will they ever be ready to make a serious and trusted contribution to the

assessment of science and scientists? As James Wilsdon, lead author of the UK report, writes in a World View on page 129, the one certainty in this debate is that the lure of metrics will only increase. Scientists should not stick their heads in the sand and pretend that the issue will go away. Rather, they should engage with metrics and work to improve the evidence base for them.

British universities now track the output of their academic staff using systems to gather details about their funding and types of output — patents, papers, citations and research grants — and to analyse institutional strengths for comparison with rival universities.

A sophisticated infrastructure has sprung up to support this activity. But it is patchy and inconsistent, with university managers often hopping between various approaches. Some, for example, have built their own internal research-information systems, and others rely on online databases of researcher outputs collected by funding agencies. There are non-profit systems that use public information, and commercially owned databases of bibliometric citations. A host of commercial benchmarking services can analyse the information. These analytical services are becoming increasingly sophisticated. They feature many different ways to group citation metrics, to cover collections of papers by individual, department, institution or journal, and to benchmark them against similar groups.

The problem is that most of these metrics tools lack transparency. At the heart of the system, databases of academic outputs and citations are not publicly accessible or auditable. And the indicators built on top of these databases can also be black boxes: the UK report notes that there are no fewer than ten major global rankings of universities, for example. Some use poorly explained scores and arbitrary weightings to underpin their league tables, and as the report says, they "assume degrees of objectivity, authority and precision that are not yet possible to achieve in practice". To some extent, metrics are used and quoted simply because other universities use them — the supply of league tables creates its own demand.

Such opacity can lead to distrust, negating the very advantage of metrics over qualitative assessment as objective, open measures of research performance. It is essential, therefore, that universities are open about the metrics that they build and use.

> **"It is essential that universities are open about the metrics that they build and use."**

Transparency is one of the hallmarks of 'responsible metrics', a term introduced by the report that covers principles such as using robust data and applying diverse indicators that account for variation by field and for multiple research types. Other principles include being humble about the limits of quantitative evaluation — which the report notes should support, rather than replace, expert assessment — and recognizing that indicators must change over time.

Although it seems legitimate to use a range of metrics to analyse research performance, their use as managerial targets can leave academics feeling 'painted by numbers' — requiring them to change their behaviour to meet often-arbitrary goals. Institutions should therefore publicly state their principles to research managers and explain why they are using particular indicators as a management tool, as the report recommends. Perhaps the most important aspect to recognize about metrics is that they can make judgements more objective — but they can also objectify those being judged. ∎

Cloud cover

Opposition to storing vast scientific data sets on cloud-computing platforms is weakening.

Earlier this week, the trade magazine *Computer Weekly* ran a short online news story about a British genomics project buying a commercial system to store its data. Genomics England, which runs the 100,000 Genomes Project, had decided to "reject" the chance to develop its own open-source system, the magazine reported.

In the past, the finer details of IT procurement were not a hot topic for researchers, and so were largely ignored by *Nature*. No longer. Just as a budding journalist cannot hope to flourish these days without a decent working knowledge of the web and multimedia skills, so young (and not-so-young) scientists must increasingly navigate the landscape of large-scale digital-information management.

As a workshop on scientific computing in Portland, Oregon, last month put it: "Computational and data-driven sciences have become the third and fourth pillar of scientific discovery in addition to experimental and theoretical sciences." In the era of big data, researchers — and journals — simply have to know their HPC (high-performance computing) from their IOPS (input/output operations per second).

'Big' barely does justice to the scale of modern scientific data. Mega, giga, tera: all are becoming increasingly familiar — then redundant — terms as the sheer colossus of research data continues to build. The Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, can generate some 25 million gigabytes of data each year — around ten times the estimated storage capacity of the human memory. Just where are we going to put it all?

A new destination has emerged in recent years: stick it in the cloud, the pervasive web-based services that will, for a fee, take your files off your hands. Late last month, the Broad Institute of MIT and Harvard, a biomedical and genomic research centre in Cambridge, Massachusetts, announced a partnership with Google Genomics to use its cloud-computing platform to store, analyse and share data. Other clouds are available, and scientists have hooked up with many of them.

In a Comment piece on page 149, several senior scientists call for this trend to accelerate. Major funding agencies, they say, should pay to place large biological data sets on the servers of the most popular cloud services — Google, Amazon and Microsoft among them. Authorized scientists would then be able to tap easily and relatively cheaply into this "global commons". The US National Institutes of Health (NIH) has cleared the way for such a move: earlier this year it lifted its 2007 ban on using cloud computing to keep and work with its own genetic database.

The NIH had been anxious about the possible threat to the privacy of those who had submitted samples. Such concerns are even more acute in Europe, where the European Commission is already engaged in an ambitious effort to crack down on how personal information is used online. (Scientists have flagged concerns that proposed new data-protection regulations could inadvertently damage clinical research; see *Nature* **522,** 391–392; 2015.) So it is reassuring that the commission has pledged to increase the access to scientific data through a continent-wide cloud-computing platform.

As we report in a News story on page 136, plans for one possible model for the European Open Science Cloud are gaining momentum, following a meeting in Geneva two weeks ago. Supporters of the project say that it would reassure academics who are reluctant to use commercial cloud services for security reasons or for fear of being tied to a particular provider. Some of the millions of gigabytes produced by CERN have already gone into a prototype system called the Helix Nebula Marketplace, involving commercial IT providers, and the lab is among those pushing for the idea to be scaled up. As a striking graph published in the Comment illustrates, the number of geneticists using cloud-based services is rising rapidly. Astronomers and other researchers are doing the same. At the very least, almost all researchers should explore the options. For more, watch this space. ∎

↻ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv

JAMES WILSDON

We need a measured approach to metrics

Quantitative indicators of research output can inform decisions but must be supported by robust analysis, argues **James Wilsdon**.

Metrics evoke a mixed reaction from the research community. A commitment to using data and evidence to inform decisions makes many of us sympathetic to, even enthusiastic about, the prospect of granular, real-time analysis of our own activities. If scientists cannot take full advantage of the possibilities of big data, then who can?

Yet we only have to look at the blunt use of metrics such as journal impact factors, *h*-indices and grant-income targets to be reminded of the pitfalls. Some of the most precious qualities of academic culture resist simple quantification, and individual indicators can struggle to do justice to the richness and plurality of our research. Too often, poorly designed evaluation criteria are distorting behaviour and determining careers. At their worst, metrics can contribute to what Rowan Williams, the former Archbishop of Canterbury, calls a "new barbarity" in our universities. Metrics hold real power: they are constitutive of values, identities and livelihoods.

Since April 2014, I have chaired an independent review of the use of research metrics for the UK government. This week, we publish the results (go.nature.com/smbaix).

They will feed into how British funding bodies will design the next round of research assessment in universities, which is used to allocate around £1.6 billion (US$2.5 billion) of funding each year. And they will be of interest to any scientist who feels the rising tide of metrics lapping at their ankles. For the research community still has the ability and opportunity — and now a serious body of evidence — to influence how this tide washes through higher education and research.

One certainty is that the lure — and so the fear — of metrics will continue. There are growing pressures to audit and evaluate public spending on higher education and research, and policy-makers want more strategic intelligence on research quality and impact. Institutions need to manage and develop their strategies for research, and at the same time compete for prestige, students, staff and resources. Meanwhile, there is a massive increase in the availability of real-time big data on research uptake, and in the capacity of tools to analyse them.

In a positive sense, wider use of quantitative indicators, and the emergence of alternative metrics for societal impact, could support the transition to a more open, accountable and outward-facing research system. Yet only a minority of the scientists we consulted supported the increased use of metrics. It is clear that across the research community, the description, production and consumption of metrics remains contested and open to misunderstanding.

Our conclusion is that metrics should support, not supplant, expert judgement. Peer review is not perfect, but it is the best form of academic governance we have, and it should remain the main basis by which to assess research papers, proposals and individuals.

Quantitative indicators can meet their potential only if they are underpinned by an open and interoperable data infrastructure. How underlying data are collected and processed — and the extent to which they remain open to interrogation — is crucial. Without the right identifiers, standards and semantics, we risk developing metrics that are not contextually robust or properly understood.

Universities, funders and publishers need to harmonize their systems of data capture. And they need to make it easier to find and assess fragmented information about research — particularly about funding. If metrics are to be reliable, and not add administrative burden, the priority for the community must be the widespread introduction of unique identifiers, such as ORCID tags, for individuals and research works.

It is tempting to boil down complex judgements to simple scores and numbers, but there is legitimate concern that some quantitative indicators can be gamed, or lead to unintended consequences. Personnel managers and recruitment or promotion panels should be explicit about the criteria they use for decisions about academic appointments and promotions. These criteria should be founded in expert judgement and may reflect both the academic quality of outputs and wider contributions to policy, industry or society.

Such decisions will sometimes be usefully guided by metrics, if the measures are relevant to the criteria in question and are used responsibly. Article-level citation metrics can be useful indicators of academic impact as long as they are interpreted in the light of disciplinary norms and with due regard to their limitations. Journal-level metrics, such as impact factors, should not be used in this way. To reduce the likelihood of abuse, publishers should stop their unhealthy emphasis on the journal impact factor as a promotional tool.

The research community needs to develop a more sophisticated and nuanced approach to metrics. (Even using the term metrics is a problem, because it implies precision and specificity. 'Indicators' is better.) Discussion is crucial, and I invite *Nature*'s readers to share good and bad uses of metrics at our new blog www.ResponsibleMetrics. org. Borrowing from the *Literary Review*'s 'Bad Sex in Fiction' award, every year we will award a 'Bad Metric' prize to the most egregious example of an inappropriate use of quantitative indicators in research management. Sadly, I imagine there will be plenty to choose from. ■

> ## THERE IS LEGITIMATE **CONCERN** THAT SOME QUANTITATIVE **INDICATORS** CAN BE **GAMED.**

James Wilsdon *is professor of science and democracy at the University of Sussex, UK, and chair of the Independent Review of the Use of Metrics in Research Assessment & Management.*
e-mail: j.wilsdon@sussex.ac.uk

CROP SCIENCE

A gene for better and longer rice

A gene that can improve the quality of rice without reducing yield has been identified by two separate teams.

Long, slender grains are considered a mark of quality for rice in many parts of the world. Xiangdong Fu of the Chinese Academy of Sciences in Beijing and his team mapped the genomes of 4,500 plants from a long-grain rice variety and zeroed in on a gene known as *Os07g0603300*. Upregulation of this gene increases cell division in the longitudinal direction and decreases it in the transverse direction. This results in a long, thin grain with very little 'chalkiness' — an undesirable opaque appearance — and no yield penalty. The gene can be repressed by a transcription factor encoded by a neighbouring gene.

Independently, Qian Qian of the Chinese Academy of Agricultural Sciences in Shenzhen, and his team discovered extra copies of *Os07g0603300* in rice varieties with these desirable traits.
*Nature Genet.* http://dx.doi.org/10.1038/ng.3352; http://dx.doi.org/10.1038/ng.3346 (2015)

PHYSICS

Tighter limits on dark matter

Atomic spectroscopy can aid the search for ultralight dark matter.

Ken Van Tilburg at Stanford University, California, and his team measured the energy emitted as atoms of the rare-earth element dysprosium transitioned between two electronic states of very similar energy over a two-year period.

They looked for fluctuations
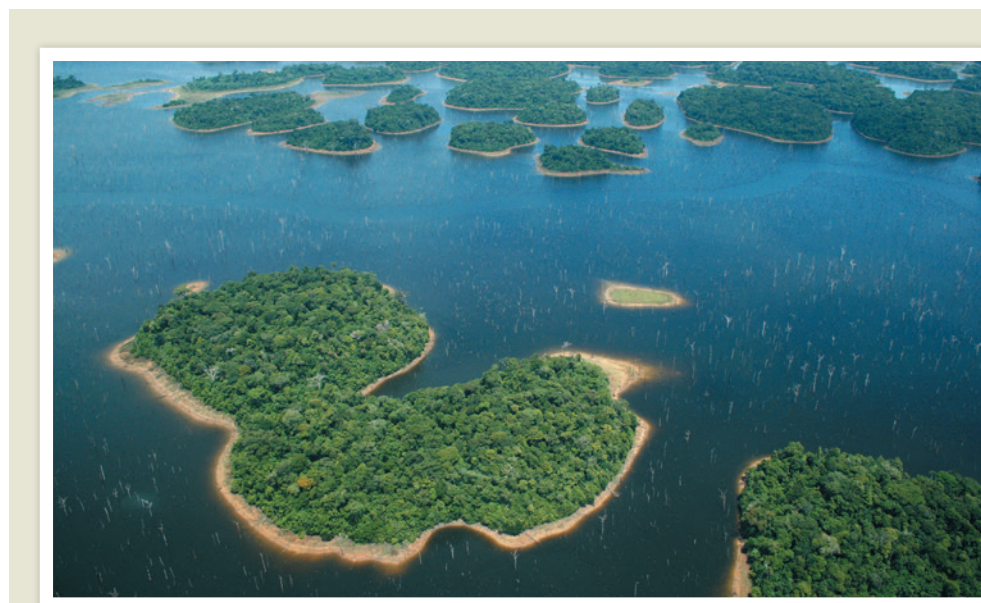
EDUARDO M. VENTICINQUE

CONSERVATION

Amazon wildlife hit by hydropower

A major hydroelectric dam in the Amazon basin has severely reduced biodiversity.

Brazil's Balbina Dam left more than 3,000 square kilometres of Amazonian forest underwater and created thousands of islands (pictured) when it was built in 1986. Maíra Benchimol and Carlos Peres from the University of East Anglia in Norwich, UK, surveyed 37 of these islands for 35 large and medium-sized mammal, bird and reptile species, using walking surveys and motion-activated cameras.

They estimate that, in the 26 years between the dam's construction and their survey, isolation has led to an overall species loss of 70% across all islands created by the dam, with smaller islands suffering the most. Just 25 of the 3,546 islands are likely to host 80% or more of the animals that they looked for. Such negative impacts have not been generally considered, and biodiversity impacts should be better assessed before future hydropower projects go ahead, the authors suggest.
PLoS ONE http://doi.org/5xh (2015)

in this energy over time, which would reveal short-term, local changes in the strength of the electromagnetic force. These could be caused by interactions with certain ultralight dark-matter particles.

No fluctuations were observed, meaning that any such dark-matter particles interacting would have to be heavier than 3×10^{-18} electronvolts or would have to interact very weakly. The results improve on previous bounds for the strength of such interactions by four orders of magnitude. If similar measurements were performed

with atomic clocks, the limits might be improved by another order of magnitude.
Phys. Rev. Lett. 115, 011802 (2015)

ANIMAL BEHAVIOUR

Flying spiders also sail on water

Spiders that use wind to carry them to new locations not only can survive a landing on water, but can also sail, even on fairly turbulent surfaces.

Many spiders exhibit 'ballooning' behaviour — they spin silken sails to travel

long distances on the wind. It had been thought that encountering water would be fatal. But Morito Hayashi at the Natural History Museum, London, and his team found that some species could survive on fresh and salt water in laboratory tests carried out at the University of Nottingham, UK, and would raise their legs or abdomens to use as sails to move across the surface. The spiders also used silk to anchor themselves in place while afloat.

This ability to control movement on water counterbalances the risks of

ballooning by helping them to survive watery landings.
BMC Evol. Biol. **15,** 118 (2015)

PLANT SCIENCE

A gene for evening scents

Petunias release their scent following the daily rhythm of a circadian-clock gene.

Takato Imaizumi of the University of Washington in Seattle and his colleagues identified a gene that they call *PhLHY* in the fragrant flower *Petunia hybrida*, which releases volatile scent molecules primarily at night.

This gene is typically expressed in the morning, dampening the expression of other genes and the production of enzymes involved in producing scent molecules.

Plants engineered to constantly express PhLHY stop producing scent molecules entirely. By contrast, plants in which this gene's expression is reduced show peak scent production around midday.
Proc. Natl Acad. Sci. USA http://doi.org/5xg (2015)

ASTRONOMY

Event pile-up may explain solar storm

A rare combination of factors might have combined to make a solar storm in March 2015 the strongest seen for a decade.

Like most such storms, this one began when the Sun spurted fast-moving plasma in an event called a coronal mass ejection. A different part of the Sun then sent out a stream

of plasma as 'solar wind'. This wind could have pushed the coronal mass ejection from behind, suggests a team led by Ryuho Kataoka at the National Institute of Polar Research in Tokyo. The whole mass could have then ploughed through space, piling up dense particles from earlier blasts of solar wind ahead of it. The Sun's magnetic field lines also happened to be oriented to drive the storm powerfully towards Earth. On hitting Earth's atmosphere, it sparked aurora (**pictured**) around the Northern Hemisphere on 17 March.
Geophys. Res. Lett. http://doi.org/5wn (2015)

CHEMISTRY

A boost for magnetic imaging

Signals from magnetic resonance imaging (MRI) can be substantially enhanced by 'hyperpolarizing' nuclear spins.

Nuclear magnetic resonance and MRI rely on powerful magnets to align the nuclear spins of protons in atoms, which then emit radio signals on returning to their normal states. These signals can be recorded to produce images or provide information on chemical composition.

Neal Kalechofsky at Millikelvin Technologies in Braintree, Massachusetts, James Kempf at Bruker Biospin Corporation in Billerica, Massachusetts, and their colleagues at these lab-equipment companies demonstrate a way to boost signals from an isotope of carbon used in medical imaging by around 1,600-fold. Their 'brute-force' approach uses low temperatures and high magnetic fields to align the spins of more atoms in a sample at 2.3 kelvin and 14 tesla than is usually possible for MRI. Samples are then ejected from the low-temperature environment, dissolved and finally transferred for imaging at room temperature and 1 tesla, providing better signals.
J. Am. Chem. Soc. http://doi.org/5x8 (2015)

Publishing delays raise hackles

While waiting for his paper to be published, Daniel Himmelstein, a PhD student in biological and medical informatics at the University of California, San Francisco, compiled the median time between acceptance and publication for 3,476 journals. He found that the wait ranged from 3 to more than 100 days.

Long delays are common, he noted on his blog. Among 16 journals in his field, *PLoS Computational Biology* was the worst, with a median wait time of 57 days. *PLoS Genetics* was not far behind. (*Nature*'s median wait was almost 48 days.)

The blog post caused a stir on social media. "Wow, @ PLOSCompBiol and @PLOSgenetics take their sweet time getting papers published," tweeted Claus Wilke, an integrative biologist at the University of Texas at Austin. David Knutson, a spokesperson for the journals' publisher the Public Library of Science, says that producing high-quality papers takes time, but that the publisher has a "laser focus" on reducing delays.

↪ NATURE.COM
For more on popular papers:
go.nature.com/slmose

VIROLOGY

Mapping viral disease vectors

The mosquitoes that carry the dengue and chikungunya viruses are more widespread than ever before, and are likely to increase their ranges.

Simon Hay at the University of Oxford, UK, and his team compiled more than 40,000 records of the occurrence of the mosquitoes *Aedes aegypti* and *Aedes albopictus*. They combined this with environmental data to map the current and possible range of these insects at a 5×5-kilometre scale.

These two *Aedes* species are found widely in all continents except Antarctica, but are still not reported in habitat that is potentially suitable for them. The team's maps could direct surveillance of these mosquitoes in understudied areas.
eLife http://doi.org/5tz (2015)

BIOMECHANICS

Seahorses benefit from square tails

The unusual square tails of seahorses both help the

animals' grasping ability and increase their toughness.

Seahorses use their bone-armour-plated tails to grip the corals and plants in which they hide, but, unlike most animal tails, the cross-section of theirs is square rather than circular. Michael Porter at Clemson University, South Carolina, and his team printed 3D articulated models of both square and circular tails and tested them under various conditions. Although the twisting ability of the cylindrical model was greater, the square prism structure provided the tail with more contact area for gripping objects and assisted the tail in relaxation, which could reduce the amount of energy a seahorse expends on grasping.

The square tail was also three times stiffer and four times stronger when compressed.
Science http://doi.org/5z9 (2015)

↪ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

PEOPLE

Researcher jailed

Dong-Pyou Han, a former biomedical scientist at Iowa State University in Ames, was sentenced on 1 July to 57 months in jail for fabricating and falsifying data in HIV vaccine trials. He was also fined US$7.2 million and will be subject to three years of supervised release after he leaves prison. His case had a higher profile than most, attracting interest from Iowa senator Charles Grassley. Han's sentence raises questions about how alleged research fraud is handled in the United States. See page 138 for more.

New ESA head

Johann-Dietrich Wörner started as director-general of the European Space Agency (ESA) on 1 July. Formerly the chairman of the German Aerospace Center (DLR), the civil engineer will serve a four-year term. He succeeds Jean-Jacques Dordain, who led the agency from 2003. Wörner plans to continue ESA's existing programmes, including the Rosetta mission, Gaia space observatory and Copernicus observation programme. He will also prepare for what he calls "Space 4.0", a phase during which space becomes a day-to-day consideration for industry and society in general.

Science panel

David King, former UK chief science adviser, is to help the European Commission to set up a new scientific-advice mechanism. On 6 July, the European research commissioner Carlos Moedas appointed King and two other experts — Dutch law scholar Rianne Letschert and former deputy prime minister of Portugal António

Joy as solar plane breaks flight record

The aeroplane *Solar Impulse 2* broke the record for the longest non-stop solar-powered solo flight on 3 July. It landed at Kalaeloa Airport in Honolulu after travelling for 4 days, 21 hours and 52 minutes and covering 7,212 kilometres. The trip was the riskiest leg of an attempt to fly around the world, starting from Abu Dhabi in March, relying exclusively on solar power. Pilot André Borschberg (right) flew the craft from Nagoya, Japan, and Bertrand Piccard (left) will fly the plane on to Phoenix, Arizona.

Vitorino — as members of a committee tasked with recommending to the commission potential candidates for a seven-strong science advisory panel that is to begin work in autumn.

Academy chief

Editor-in-chief of the Science group of journals, Marcia McNutt, was nominated on 6 July to stand for election as president of the US National Academy of Sciences (NAS). If elected, as expected, she will become the first woman to head the organization since its inception in 1863. McNutt, a geophysicist, was appointed the first female editor-in-chief of *Science* — published by the American Association for the Advancement of Science — in 2013, where she will continue until the NAS's current president, Ralph Cicerone, ends his second term on 1 July 2016.

EVENTS

Pluto probe scare

NASA's New Horizons probe stopped recording science data on 4 July, ten days before it is to fly past Pluto in the first-ever visit to that dwarf planet. Mission controllers lost communication with the probe for 81 minutes, but recovered it completely the following day. Roughly 30 science observations were lost during the glitch, which happened when the onboard processor tried to simultaneously compress data that had already been gathered and write a sequence of future flight commands into its flash memory. This caused the probe's main computer to enter a 'safe mode'. See go.nature.com/dckjgk for more.

Liberia Ebola

Authorities in Liberia are scrambling to find out how a 17-year-old boy became infected with the Ebola virus and died, becoming the first case since the country was declared free of the disease on 9 May. The World Health Organization reported the case on 3 July. The boy died on 28 June in Margibi county close to the capital Monrovia, far from the borders with Sierra Leone and

Guinea — where the epidemic continues at low levels. Authorities are trying to trace the boy's contacts. Of the 200 identified so far, two have tested positive for the virus and have been isolated.

Farming outlook

All is calm on the world food-supply and price front, at least for now, say the Organisation for Economic Cooperation and Development (OECD) and the Food and Agriculture Organization of the United Nations (FAO). Their latest annual *Agricultural Outlook*, released in Paris on 1 July, predicts a slide in farm-product prices over the next decade owing to higher crop yields and productivity, and slower growth in demand. OECD Secretary-General Angel Gurría said "The outlook for global agriculture is calmer than it has been in recent years." But he added that price spikes in the coming years cannot be ruled out.

BP oil settlement

Oil company BP would pay US$18.7 billion over 18 years to settle civil lawsuits related to the 2010 Deepwater Horizon oil spill (**pictured**), under a tentative settlement with US state and federal governments. The deal, announced on 2 July, would be the largest settlement with a corporation in US history, according to

the US Department of Justice. Before the agreement can be finalized, it must undergo a public comment period and review by a federal court. See go.nature.com/xsus9t for more.

Telescope dreams

NASA and its international partners should build a space telescope five times the size of the current Hubble Space Telescope, an influential group of US astronomers says in a 6 July report. This 12-metre 'High-Definition Space Telescope' would be the first true Hubble replacement; the James Webb Space Telescope, NASA's next big observatory launching in 2018, operates in infrared light and not the visible and ultraviolet wavelengths that Hubble uses. The report, from the Association of Universities

for Research in Astronomy in Washington DC, does not specify a cost or time frame for building the telescope. See go.nature.com/vskh3s for more.

Greek scientists

The economic crisis in Greece is hitting researchers hard, with Greek scientists losing access to some digital journals. The Internet portal that provides many Greek universities and research institutes with access to electronic journals from 27 publishers suspended many of its services on 1 July because the government has not provided funds to keep it going. The Hellenic Academic Libraries Link has nearly shut down many times over the past decade. But now with the threat of state bankruptcy, scientists are not expecting rescue funds. See go.nature.com/vgc5wj for more.

Vaccine push

California governor Jerry Brown signed into law on 30 June a bill that mandates vaccinations for all children attending public schools. Parents can no longer choose not to vaccinate these children for religious or ideological reasons. Exemptions would be granted only for medical reasons. The move was sparked by

a measles outbreak last December that could be partly attributed to low vaccination rates, and comes in the same week as the United States announced its first measles death in 12 years on 2 July. California is only the third US state to ban vaccine exemptions that are based on personal and religious beliefs.

Cystic-fibrosis drug

US regulators have approved a drug to treat the most common form of cystic fibrosis. On 2 July, the US Food and Drug Administration announced that it had approved Orkambi (lumacaftor and ivacaftor) for people with cystic fibrosis who have two copies of a mutation called F508del in the CFTR protein. Orkambi is made by Vertex Pharmaceuticals of Boston, Massachusetts, a company that has pioneered cystic-fibrosis treatments that target the cause of the disease (see *Nature* **482**, 145; 2012), rather than just the symptoms.

↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news
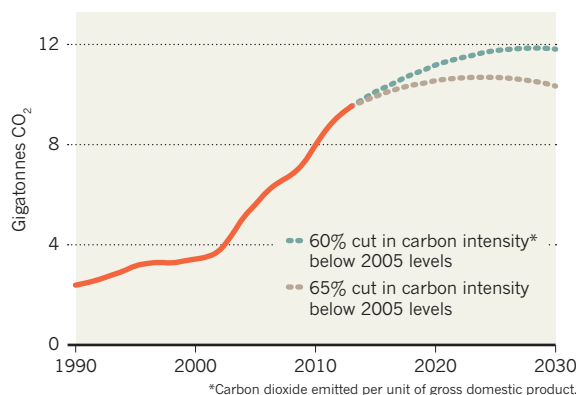
TREND WATCH

China submitted its pledge to cut carbon emissions to the United Nations on 30 June. The country pledged to boost renewable energies and to reduce the amount of carbon dioxide emitted per unit of gross domestic product (carbon intensity) to 60–65% below 2005 levels. An analysis by GWG Energy in London suggests that China could level off its emissions before 2030 if it meets the carbon-intensity target, depending on how fast its economy grows. See go.nature.com/3bkybj for more.

ON THE LEVEL

China's latest pledge to reduce its carbon dioxide emissions could allow the country's carbon output to peak by 2030.
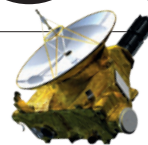

60% cut in carbon intensity* below 2005 levels
65% cut in carbon intensity below 2005 levels

*Carbon dioxide emitted per unit of gross domestic product.

Meteorologists cruise country roads at night to put their instruments in the path of heavy weather.

METEOROLOGY

Night–time storm chasers stalk their prey on US Plains

Violent nocturnal thunderstorms are hard to explain and even harder to forecast.

BY ALEXANDRA WITZE, STRONGHURST, ILLINOIS

Sheets of rain pummel the Illinois country-side as Jacey Wipf and Kyle Morganti haul a 60-kilogram weather station out of their pick-up truck. They rotate the metal cylinder on the roadside gravel, level it and step back to photograph its surroundings. Then they dash for the safety of their truck as bolts of lightning strike uncomfortably close, illuminating the pitch-black June night.

As technicians with the Center for Severe Weather Research in Boulder, Colorado, Wipf and Morganti are accustomed to this sort of extreme fieldwork. They are two in an army of researchers who have descended on the US Great Plains this summer for a massive research programme that ends on 15 July. The 45-day, US$13.5-million Plains Elevated Convection At Night (PECAN) project aims to unravel the mystery of how thunderstorms form and evolve at night, long after the solar heating that fuels daytime thunderstorms has vanished.

These night-time storms bring hail, flash floods and strong winds that can damage homes and cars. Because they occur in the dark, even experienced weather-watchers cannot detect their development. And they continue to elude nearly all attempts at forecasting.

"We really cannot predict, even on a 12-hour notice, where these storms are going to be," says Bart Geerts, a PECAN principal investigator and an atmospheric scientist at the University of Wyoming in Laramie.

Understanding night-time thunderstorms could help to improve forecasts of dangerous weather events on the Great Plains, he says. The research could also apply to other parts of the world that have meteorologically similar storms, such as those on the plains of eastern South America. A project to study similar thunderstorms in Argentina is slated for 2017. These storms start during day and night in the province of Mendoza, near the Andes foot-hills, where sudden hailstorms can wipe out economically important vineyards, says Jorge Rubén Santos, an atmospheric scientist at the National University of Cuyo in Mendoza.

All the textbook theories to explain thunder-storms have been developed with reference to daytime conditions, when heat rises from the ground to produce a well-mixed layer of air that feeds burgeoning storms above. PECAN is ▶

testing alternative ideas to explain how things might be different at night, when a stable layer of cool air typically prevents warm air from rising and churning to generate storms.

One idea involves a fast-moving ribbon of air called a low-level jet, which can form when air over higher elevations cools relative to that at lower elevations, setting up a pressure gradient. Computer simulations suggest that these jets can lift moist air above the stable layer, where they can feed storms (A. J. French and M. D. Parker *J. Atmos. Sci.* **67,** 3384–3408; 2010).

"But sometimes there are just nights when you have no obvious forcing like a low-level jet," says Rita Roberts, an atmospheric scientist at the National Center for Atmospheric Research in Boulder. Other atmospheric patterns may be at play, such as wave-like structures called bores that PECAN is also hunting this summer.

Project storm-chasers have had only mixed success, observing lots of low-level jets but not as many full thunderstorm complexes as they would have liked. "It's been a frustrating year," says Matthew Parker, an atmospheric scientist at North Carolina State University in Raleigh.

Each day, the team decides where to deploy an armada of trucks, vans and aeroplanes laden with instruments including radar, radiosondes and balloons. The scientists fan out ahead of where they think the storms will move, and hope to intercept them as they sweep through.

"If we could forecast them precisely, we wouldn't need to be out here."

"We have to wait for nature to provide us with storms of different types," says Joshua Wurman, president of the Center for Severe Weather Research.

Since PECAN began on 1 June, Wipf and Morganti have clocked up long hours collecting meteorological data ahead of approaching storms. This means a lot of driving along country roads in the dark and the rain — not exactly the glamorous stereotype of storm-chasing. "Everybody always thinks it's just like *Twister*," Wipf says. "It's not."

On 24 June, they find themselves in western Illinois, swerving to avoid low-hanging trees that could take out the towering measurement mast fixed to the front of their truck. "Tree tree tree!" Wipf shouts, just before Morganti swings the wheel yet again.

At 11:29 p.m., Wipf gets a text ordering them to deploy five stations along the side of the road, spacing them every 2 kilometres to gather data on temperature, relative humidity, wind speed and pressure. They wrestle two stations out of the truck before lightning begins hitting too close and they are forced to stop for the night.

In the end, it doesn't really matter that the stations are not up and running. The worst storms pass about 80 kilometres west of the pair, because PECAN forecasters have failed once again to predict how the night's events will unfold.

"If we could forecast them precisely," says Wurman wistfully, "we wouldn't need to be out here." ■

Europe sets its sights on the cloud

Three large labs hope to create a giant public–private computing network.

BY ELIZABETH GIBNEY

From astronomy to genomics, scientists are increasingly storing and studying their data sets on shared remote 'cloud' computing servers, accessed through the Internet. Three of Europe's biggest research labs now want to help academics by working with commercial firms to create a continent-wide cloud-computing portal — and they are hoping to get backing from the European Commission.

Many researchers find cloud computing to be more flexible and efficient than buying expensive hardware — they can rent servers from firms such as Amazon and Google when they need a burst of power for an intensive computation, for example (see *Nature* **522,** 115–116; 2015). Despite the advantages, some academics are concerned about security and reliability when storing their data on outside servers, says Bob Jones, a computer scientist at CERN, Europe's particle-physics lab near Geneva, Switzerland.

Jones thinks that a single portal combining offerings from commercial providers and publicly funded infrastructure could solve some of these problems, and ultimately increase access to key data sets. Since 2012, CERN — with the European Space Agency and the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany — has been developing a test-bed system called the Helix Nebula. Run for two years with funding from the European Commission, and coordinated by Jones, the initiative has since evolved into a portal involving 30 different cloud providers, known as the Helix Nebula Marketplace (HNX). CERN has simulated particle collisions on the platform, and EMBL has used it to analyse genetic sequences, including some moved from Amazon's cloud, says Rupert Lück, EMBL's head of IT services.

Ambitions to expand were bolstered when, in May, the European Commission announced plans to fund a Europe-wide 'research cloud'. "The commission likes the idea of open science," Jones said on 26 June at a meeting in Geneva to discuss a European Open Science Cloud. "What we have to do now is take that enthusiasm from the public sector, the private sector and European institutions, and put it in place."

The commission is not specifically backing Jones' plan: it will launch its call for proposals in 2016 and says there are "a range of possibilities for business models". It wants a virtual platform to host data and encourage their analysis and reuse across disciplines and borders. Climate and satellite data, for instance, "represent a goldmine for research, innovation and new business opportunities", says the commission.

A European cloud for researchers built around the HNX would be a single gateway through which users could access cloud services and open research data from existing public infrastructure — for example, through

MORE ONLINE

Q&A

How will BP's deal help to restore the Gulf of Mexico?
go.nature.com/hiovyl

MORE NEWS

● Gene responsible for tastier rice identified go.nature.com/uwebbb
● Ultrasound scanners read fingerprints go.nature.com/8vxtg2
● Seafaring spiders use silk as sail go.nature.com/nofbfx

NATURE PODCAST

nature podcast

Unicellular eyes, ethical robots and sex-morphing lizards
nature.com/nature/podcast

the European Grid Infrastructure Federated Cloud, a network of largely publicly funded cloud services such as the Supercomputing Centre of Galicia in Spain — and through companies, such as Cloudwatt, a provider based in Paris. A pilot platform would start relatively small, with the computing equivalent of 100 million hours of processor time and some 10 petabytes of storage (1 petabyte is 10^{15} bytes). The network would need to expand to 20 times this size to serve the whole of Europe, says Jones.

An advantage of such a system is that all data would be stored, protecting them if a provider were to stop operating, says Jones. And the system's standard terms would make it quicker and easier for researchers to sign up to and access, he says. "The most valuable thing for researchers is their data. If we're going to convince researchers to trust cloud services, we really do need this hybrid model." A federated European cloud could also deal with restrictions that require sensitive data to be analysed in its country of origin, says Lück.

In the United States, researchers and funders are also thinking about how to increase access to data stored on clouds variously funded by the US National Science Foundation, individual institutions and companies, says David Lifka, director of the Cornell University Center for Advanced Computing in Ithaca, New York, which runs a service called Red Cloud. "Sharing cloud capacity is the next logical step," he says. But creating a system that is fair and does not constrain users is not easy, he adds.

US computer giants Google, Amazon and Microsoft are notably absent from the HNX. Mark Skilton, who studies information systems at the University of Warwick, UK, suggests that the focus on European companies may reflect the commission's desire to boost homegrown providers. "The issue is whether this will suffer for the lack of Amazon and Google scaling," he says. Some researchers see the likes of Amazon and Google as a route to open data. Writing in *Nature* this week, genomics researchers call on funding agencies to expand access to major data sets by paying to place them in popular cloud services (see page 149).

The biggest barrier to cloud computing for small labs is the cost of accessing high-quality cloud resources, says Skilton. If the negotiating power of a European initiative can bring costs down, many could benefit, he says. But it is unclear whether commercial providers will play ball, says Lifka. Although firms often give trial periods for free, "from my experience, their price is their price", he says. Getting everyone — especially commercial partners — to work under the same governance system and according to the same conditions will be an organizational challenge, says Skilton. ∎

The bright-line brown eye (*Lacanobia oleracea*) is just one of many potential tomato residents.

BOTANY

Plant dwellers take the limelight

Researchers seek holistic view of botanic ecosystems.

BY HEIDI LEDFORD

A plant may be rooted in place, but it is never lonely. There are bacteria in, on and near it, munching away on their host, on each other, on compounds in the soil. Amoebae dine on bacteria, nematodes feast on roots, insects devour fruit — with consequences for the chemistry of the soil, the taste of a leaf or the productivity of a crop.

From 30 June to 2 July, more than 200 researchers gathered in Washington DC for the first meeting of the Phytobiomes Initiative, an ambitious proposal to catalogue and characterize a plant's most intimate associates and their impact on agriculture. By the end of the year, attendees hope to carve out a project that will apply this knowledge in ways that will appeal to funders in industry and government.

"We want to get more money," says plant pathologist Linda Kinkel at the University of Minnesota in St Paul. "But beyond that, let's just all try to talk the same language and come up with some shared goals."

The effects of microbes and insects on plant health have often been studied in pairs — one microbe and one plant. But advances in genetic sequencing have opened up ways to survey entire microbial communities. Meanwhile, engineers and computational biologists have developed better ways to manage large data sets, merge disparate recordings into cohesive models and rapidly collect information on the physiology of every plant in a field. "Historically, we haven't had the capacity to look at this as a system," says plant pathologist Jan Leach at Colorado State University in Fort Collins. "Now we need to begin to integrate not just the data about the plant and the plant's environment, but all the biological components in that system."

Leach coined the term phytobiome in 2013, at a retreat about food security. She defines ▶

▶ the phytobiome broadly, to encompass microbes, insects, nematodes and plants as well as the abiotic factors that influence all these.

Since then, she has visited companies, funding agencies and universities to call for a unifying phytobiomes initiative. She has teamed up with Kellye Eversole, a consultant based in Bethesda, Maryland, and the co-owner of a small family farm in Oklahoma, who has experience working on large agricultural genomics projects, including the US National Plant Genome Initiative. That initiative was launched in 1998 and continues to crank out databases and other tools for analysing plant genomes.

Leach hopes that the Phytobiomes Initiative will leave a similar legacy, but she is mindful that federal funding has tightened considerably since 1998. Still, she notes that the project can build on several emerging trends in agriculture. Industry has shown renewed interest in boosting plant growth by manipulating associated microbes (*Nature* **504**, 199; 2013). Companies and farmers are also investing in 'precision agriculture', which uses high-tech monitors to track conditions in a field or even around individual plants, allowing farmers to water and fertilize in exactly the right places.

HIGH-TECH FUTURE

Eversole foresees a day when tractors will carry dipstick-like gauges that provide a snapshot of the microbial community in the soil. Data from the Phytobiomes Initiative would then help farmers to manipulate that community to their advantage, she says.

But first, the initiative needs to standardize protocols and metrics, the meeting's attendees determined. Kinkel says that efforts are likely to focus initially on cataloguing microbes and insects and their interactions with different crops and habitats. "We're where plant biologists were 150 years ago," she says. "We're still trying to inventory things."

Work has already begun along these lines: for example, a group at the International Rice Research Institute in Los Baños in the Philippines is fishing for microbial DNA in data discarded from an effort to sequence the rice genome. The goal is to determine which microbes prefer which strains of the crop.

Kinkel, meanwhile, has begun experimenting with manipulating carbon levels in the soil to alter the microbial population, with the aim of improving plant productivity. "If we can understand better who lives on and within plants, we have the potential to manage them to have healthier, more resilient plants," she says.

Projects such as these would move faster under an organized, cohesive framework, says Sarah Lebeis, a microbiologist at the University of Tennessee in Knoxville who is studying how plants manipulate microbial communities by secreting antibiotics into the soil. "Right now we're working as individuals," she says. "Having an initiative will give us focus and hopefully we'll progress further, faster, better." ■

Dong-Pyou Han (centre) confessed to fabricating and falsifying data on an HIV vaccine.

RESEARCH MISCONDUCT

Uneven response to scientific fraud

The case of jailed US vaccine researcher Dong-Pyou Han shows up inconsistent nature of penalties.

BY SARA REARDON

Rare is the scientist who serves time on charges of research misconduct. But on 1 July, Dong-Pyou Han, a former biomedical scientist at Iowa State University in Ames, was sentenced to 57 months in prison for fabricating and falsifying data in HIV vaccine trials. Han has also been fined US\$7.2 million and will be subject to three years of supervised release after he leaves prison.

His case had a higher profile than most, attracting interest from a powerful US senator. Han's harsh sentence raises questions about how alleged research fraud is handled in the United States, from decisions about whether to prosecute to the types of punishment imposed by grant-making agencies.

Han was forced to resign from Iowa State in 2013, after the university concluded that he had falsified the results of several vaccine experiments supported by grants from the US National Institutes of Health (NIH). In some cases, Han spiked rabbit blood samples with human HIV antibodies so that the vaccine seemed to have caused the animals to develop immunity to the virus.

In a confessional letter sent to the university just before its investigation concluded, Han said that he began the subterfuge to cover up a sample mix-up that he had made years before.

The US Office of Research Integrity (ORI), which oversees investigations into alleged misconduct involving NIH funds, barred Han from receiving federal grants for three years — the maximum penalty that it generally imposes on junior investigators. The case probably would have ended there had it not drawn the attention of Senator Charles Grassley (Republican, Iowa), who has a history of investigating misconduct in the biomedical sciences.

"This seems like a very light penalty for a doctor who purposely tampered with a research trial and directly caused millions of taxpayer dollars to be wasted on fraudulent studies," Grassley wrote in a February 2014 letter to the ORI. The office can issue lifetime funding bans,

This story is the first in an occasional series on research misconduct in the United States.

but former ORI officials say that such punishment is reserved for especially egregious cases, such as those in which human subjects could have been endangered.

In June of that year, after extensive media coverage of the case and of Grassley's reaction to it, the federal prosecutor in Des Moines, Iowa, pressed charges against Han. The scientist was arrested and his case brought before a grand jury. In February 2015, he pleaded guilty to two felony charges of making false statements to obtain NIH research grants.

Alan Price, a former associate director of investigative oversight at the ORI, says that criminal prosecution is unusual for a "medium-level" fraud case such as Han's. "In most cases, I don't think it would have been done. But Senator Grassley cares deeply about these issues and wanted to make that point."

The case has raised some concern among experts in scientific misconduct. The very few researchers who face criminal charges are not necessarily those who have caused the most harm to other scientists' careers, or to science generally. "We're so preoccupied with major cases and so subject to policy pressure, we've lost sight of the larger picture," says Nicholas Steneck, an expert in research integrity at the University of Michigan in Ann Arbor.

Grassley seems to agree — telling the Senate in July, "I worry that other cases may go unnoticed or unaddressed if there isn't a public outcry". He argues that lawmakers would not need to involve themselves in such matters if some government agencies that oversee research grants could levy harsher penalties and had more capacity to investigate alleged fraud.

Most US funding agencies have an inspector-general who investigates potential misconduct and fraud. These officials can withdraw grant money and impose prohibitions on receiving government funds, and often refer cases for criminal prosecution.

But the Department of Health and Human Services (HHS), which includes the NIH and the ORI, separates these powers. The ORI cannot directly investigate suspected fraud or misconduct; it is limited to overseeing probes by the institutions that employ the researchers suspected of wrongdoing. In cases where evidence of misconduct or fraud is found, the ORI can impose funding bans or refer potential criminal cases to the Department of Justice or the HHS inspector-general.

The HHS inspector-general can initiate investigations of suspected research fraud or misconduct, but is often preoccupied with other matters, such as health-insurance

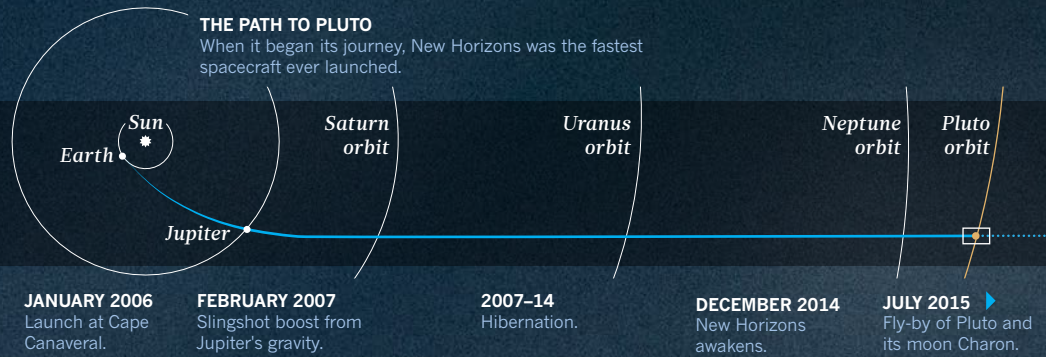> *"It's questionable how much more is to be gained by jail time."*

fraud. And it cannot impose funding bans or other penalties. The NIH and the ORI told *Nature* that they do not even track how many recipients of NIH grants have faced criminal prosecution.

By contrast, the inspector-general for the National Science Foundation has sole oversight of that agency's misconduct investigations, and is involved in several criminal prosecutions each year. Most of these concern researchers suspected of misusing grant money or of using plagiarized or falsified data to obtain funds, as Han did.

But David Wright, a former ORI director, says that the benefit of criminal prosecution is unclear. Formally barring a researcher from receiving federal funds is usually a professional death sentence, even if the ban is short, he adds. "It's questionable how much more is to be gained by jail time."

In reality, however, no one knows the general fate of scientists subject to funding bans, or whether the risk of such punishment deters people from committing misconduct. Price says that he and others at the ORI once tried to conduct a formal, anonymous survey of these researchers to understand how their careers had been affected. But the White House shut the project down, saying that it cost too much and that people were unlikely to respond. ∎

THE PATH TO PLUTO
When it began its journey, New Horizons was the fastest spacecraft ever launched.

Sun
Earth
Jupiter
Saturn orbit
Uranus orbit
Neptune orbit
Pluto orbit

JANUARY 2006
Launch at Cape Canaveral.

FEBRUARY 2007
Slingshot boost from Jupiter's gravity.

2007–14
Hibernation.

DECEMBER 2014
New Horizons awakens.

JULY 2015 ▶
Fly-by of Pluto and its moon Charon.

24 HOURS OF PLUTO
SPEEDING PAST AN ICE WORLD AT THE FRINGES OF THE SOLAR SYSTEM

BY ALEXANDRA WITZE
DESIGN BY JASIEK KRZYSZTOFIAK

On 14 July, after a journey of nine and a half years and some 5 billion kilometres, NASA's New Horizons spacecraft will visit the frigid frontier of the Solar System: Pluto. It will be a fast and furious meeting — the spacecraft will whiz past at nearly 50,000 kilometres per hour, collecting photographs and scientific data on Pluto's surface, atmosphere and environment during the 24-hour event. No mission has ever visited Pluto or any of the other ice worlds that make up the Kuiper belt, the swarm of small and frosty bodies that orbit mostly beyond Neptune. With its huge moon Charon, Pluto also constitutes the Solar System's only known binary system.

THE DWARF PLANET

SURFACE
Pluto is covered with several types of ice, including methane, nitrogen and carbon monoxide. Its reddish surface is one of the most strongly mottled in the Solar System, and New Horizons should reveal the identities of these light and dark patches. Its closest analogue in the Solar System may be Neptune's icy moon Triton, which is thought to have been captured from the Kuiper belt.

ATMOSPHERE
Pluto has a thin atmosphere generated by ices sublimating from its surface. Since its discovery in 1988, the atmosphere has mysteriously expanded — even though Pluto is getting farther from the Sun.

▶ NATURE.COM
Visit **www.nature.com/pluto** for more on Pluto.

Ocean?
Rocky core?
Icy surface and mantle
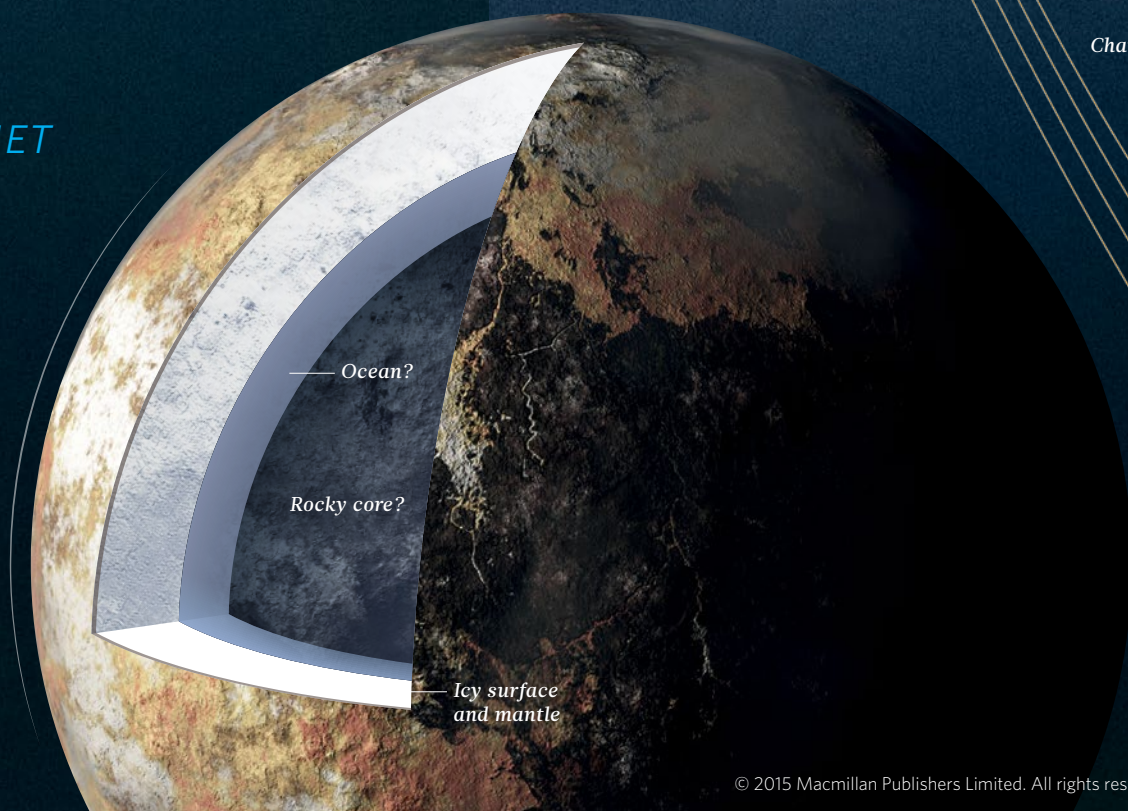
THE FLY-BY

Up to and including 12 JULY
New Horizons will map the surface and study the atmosphere, looking for clouds and haze on Pluto, as well as rings and moons beyond the five known (Charon, Styx, Nix, Kerberos and Hydra).

13 JULY
Limited initial observations will be sent back to Earth in case the spacecraft does not survive the encounter.

14 JULY
New Horizons will remain radio silent for much of the day so that it can concentrate on gathering data at Pluto and Charon. It will collect colour Images of Pluto at a resolution of 0.5 kilometres per pixel, and black-and-white ones (in a narrow band across the dwarf planet's centre) at resolutions as high as 100 metres per pixel.

15 JULY
Close-up images of Pluto and Charon, along with scientific data, will start to be sent to Earth over a 26-month period. New Horizons' transmission rate is limited by its communications time with NASA's Deep Space Network and the sheer quantity of data that it will collect during the intense, close encounter. The highest-resolution images of Pluto that will be available from the encounter will be transmitted on 15 July, with those for Charon following the day after.

14 JULY

7:50 a.m. Eastern Daylight Time
Closest approach to Pluto, at 12,500 kilometres. Images taken in both visible and near-infrared wavelengths.

8:04 a.m.
Closest approach to Charon, at 28,800 kilometres. Because this is more than twice the distance of the closest approach to Pluto, the best pictures of Charon will be roughly twice as coarse as those of Pluto.

8:51 a.m.
Passes through Pluto's shadow, allowing it to probe Pluto's atmosphere.

10:18 a.m.
Passes through Charon's shadow, allowing it to search for an atmosphere on Charon.

9:02 p.m.
Mission team on Earth should receive a preprogrammed 'phone home' signal which, if all went well, will indicate the spacecraft survived the encounter.

Nix
New Horizons spacecraft
Pluto
Styx
Charon
Kerberos
Hydra

THE MOONS

FORMATION
Early in the Solar System's history, a proto-Charon probably walloped into a proto-Pluto, sending debris cascading out into space. Much of that may have condensed to form Pluto's four smaller moons.

Proto-Charon
Proto-Pluto

BINARY SYSTEM
Pluto and Charon are locked in an intricate orbital dance. Because Charon is so large relative to Pluto — at one-eighth its mass — the two actually orbit a mutual centre of gravity that is located in space. They also both rotate on their axes once every 6.4 Earth days. Analyses of the shapes of Pluto and Charon could reveal whether one or both of them ever harboured an underground ocean, kept liquid by subterranean heat.

THE SMALLER MOONS
Nix and Hydra tumble chaotically on their axes, but Nix, Styx and Hydra are locked in an orbital resonance that has them travelling around Pluto in synchrony. Kerberos is surprisingly dark in colour, possibly reflecting a piece of the original impactor that formed the Pluto–Charon system. Of the small known moons, New Horizons will get the best view of Nix. It may also discover more moons, or dust rings, somewhere in the system.

Charon
Pluto

Charon
Styx Kerberos Nix Hydra
Pluto
Earth's Moon

BARRY GUTIERREZ

THE QUAKE HUNTERS

Meet the seismologists who work around the clock to pinpoint major earthquakes anywhere on Earth.

BY ALEXANDRA WITZE

At 17 minutes past midnight on Saturday 25 April, Rob Sanders's computer started chiming with alerts. On his screen, squiggly recordings poured in from seismometers in Tibet, Afghanistan and nearby areas that were feeling the first vibrations from a tremendous earthquake.

Sanders was part way through his shift as an on-duty seismologist at the US Geological Survey's National Earthquake Information Center (NEIC) in Golden, Colorado. It was his job to work out what was happening — and fast. Within 30 seconds, he began analysing the seismic data and realized it was time to call his boss.

When the phone rang, Paul Earle was dozing in the room of his four-year-old son, where he had nodded off earlier that evening. Earle rolled out of bed and logged onto his home computer. As chief of 24/7 operations at the NEIC, Earle knew that time was short. For any major earthquake in the world, the US Geological Survey (USGS) is committed to publishing the shock's magnitude and location online within 20 minutes. The team also puts out rapid estimates for how many people may have been hurt. Various nations issue alerts for quakes in their vicinity,

Seismologists at the National Earthquake Information Centre are on duty 24/7 to monitor quake activity.

but Earle's crew is the only one that analyses tremors around the globe.

The NEIC information helps governments and humanitarian groups to decide how to respond in times of crisis. It determines whether search-and-rescue teams pack their bags, and whether financial markets begin responding to a catastrophic natural disaster. When minutes count, hundreds of key responders — from the White House to the United Nations — rely on the NEIC team to tell them exactly how bad an earthquake was. On 25 April, the work that began on Sanders's screen ended up with the US government dispatching a response team to the quake's epicentre in Nepal within hours.

The NEIC seismologists do not always get it right. Sometimes, deceived by the rawness of the data, they put out an alert containing the wrong quake location or size, before quickly retracting the information. But they are continually refining their techniques to speed up response times while maintaining accuracy. "Being reliable is more important than pure speed," says Earle.

THE NIGHT SHIFT

The NEIC takes up the fifth floor of a blocky building on the campus of the Colorado School of Mines in Golden, not far from the original Coors brewery and bronze sculptures of the miners who shaped this region of Colorado. A decade ago, television satellite trucks regularly clogged the car park after any large earthquake. Now, most of the journalists stay at home — they can get information from the centre faster over the Internet.

Computer monitors have replaced the slowly rotating paper drums that once displayed the vibrations measured at seismic stations around the world. But the centre has kept one relic on display: a large wooden globe that often appeared in television reports. Patches of its coloured surface are worn away from decades of seismologists jabbing their fingers at earthquake locations. Southern California has basically disappeared. So has Japan.

Established in 1966, the NEIC originally operated during normal business hours, with seismologists on call at other times. But in 2004, a magnitude-9.1 earthquake hit Sumatra, triggering a ruinous tsunami that killed almost a quarter of a million people around the Indian Ocean. In an effort to improve its response times in major disasters, the earthquake centre moved to operating around the clock. Fourteen seismologists now cover three shifts, with at least two people on duty at any given time (coordinating their toilet and meal breaks).

The NEIC analyses more than 20,000 earthquakes a year, everything from imperceptible ones in California to the monsters that occasionally shake the globe. It reports on any earthquake of magnitude 5 or greater worldwide, and down to magnitude 3 in parts of the United States.

On 25 April, the only earthquake that mattered began beneath Nepal. The jolt started 15 kilometres underground, on the huge Himalayan fault where the tectonic plate carrying India rams into Asia. At 11:56 a.m. local time (11 minutes past midnight in Colorado), the stress of that geological collision ruptured a 120-kilometre-long segment of Earth's crust beneath the Nepalese district of Gorkha. Waves of seismic energy raced outwards in all directions.

Within 16 seconds they reached Kathmandu, almost 80 kilometres to the southeast, and began toppling thousands of buildings. Just over a minute later they passed Lhasa, 600 kilometres northeast of the epicentre, and shook seismometers bolted into granite in a hillside tunnel. Those machines, part of the Global Seismographic Network, immediately relayed their data to the NEIC.

At the Colorado centre, an alert dinged and a window popped up on Sanders's screen, which filled with information from stations around Asia. Sanders started sorting through the data, choosing the best seismic records to include in his analysis.

A second seismologist on duty that night called and woke Earle, who began to work on the seismic data from home. As the minutes ticked away, the three of them faced a crucial task — deciding on the quake's

magnitude. The USGS measures eight types of magnitude, each of which conveys different information about the strength of an earthquake's vibrations and the amount of energy it releases. Certain magnitude scales are most accurate for smaller quakes, whereas others are better at describing long-lasting, larger shocks.

At 12:29:42 a.m. — 18 minutes and 16 seconds after the earthquake began — the NEIC released its first answer. Location: 77 kilometres northwest of Kathmandu. Size: 7.5 on the moment magnitude scale. This particular scale relies on computer modelling of a certain type of seismic wave, and Earle chose it because of a gut feeling for what he thought would represent the most meaningful magnitude.

But as is often the case with large quakes, the first official magnitude was not the last. The team had only just started its analyses. Earle called and woke up two more colleagues — Harley Benz and Gavin Hayes — then jogged the two blocks from his home into work. Even as news agencies

"THAT'S WHEN WE KNEW IT WAS GOING TO BE DEADLY."

began broadcasting alerts of a magnitude-7.5 earthquake in Nepal, the NEIC researchers were sifting through fresh data.

From his home, Hayes ran a separate set of model calculations, which use data on longer-period seismic waves that arrive at stations later but are more appropriate for the world's largest quakes. At 1:04 a.m., on the basis of this 'W-phase' analysis, the NEIC updated the Nepal quake's magnitude to 7.9.

"None of those numbers are wrong," says Earle. "They're all right for that particular magnitude scale." (Three hours later, the centre would announce a final magnitude of 7.8, also based on the W-phase approach but incorporating more-detailed modelling with newer data.)

Even as Earle was wrestling with the quake's magnitude, he called NEIC seismologist David Wald, who happened to be awake. Wald runs a set of programs that take the initial magnitudes and estimate possible fatalities and economic losses. The system, called PAGER (Prompt Assessment of Global Earthquakes for Response), relies on databases of where people live, the types of building in the region of an earthquake and how many people had been killed in similar quakes in the area before.

If a quake is big enough, PAGER sends out alerts automatically. At 12:34 a.m., the system used the initial magnitude of 7.5 to predict between 100 and 1,000 deaths, and damages between US$10 million and $100 million. That ranked it an 'orange', the second-highest alert on the PAGER colour-coded system. "That's when we knew it was going to be deadly," Wald says.

As the minutes crept by, aftershocks kept pummelling Kathmandu. PAGER automatically updated three more times at the orange level, the last at 2:16 a.m.. Then Wald took some data on how much the ground had moved and how widespread the aftershocks were, and manually fed the fresh information into PAGER. The alert immediately escalated to red, estimating between 1,000 and 10,000 deaths. It was 4:14 a.m..

GLOBAL RESPONSE

In Washington DC, Gari Mayberry's mobile phone woke her up with the first NEIC alert. Mayberry, a USGS volcanologist, advises the US Agency for International Development on natural hazards. The agency funded PAGER's development, precisely to simplify split-second decisions after earthquakes. "Do I need to call my boss at 3 a.m.?" asks Mayberry. "That's what people want to know."

For Nepal, the answer was yes. As the Colorado team released its analyses, Mayberry quickly fed information to her bosses, who help to coordinate search-and-rescue teams for international disasters. In such situations, she says, every minute counts. Within hours, the US government had a team on the way to Nepal.

Paul Earle and the team at the earthquake centre issue alerts for major quakes within 20 minutes.

in Alaska and Japan. An on-duty seismologist had caught the problem, but the software that distributes the alerts had not responded to the correction.

Cutting back on false alerts while making sure that the real ones get out in time takes a nuanced mix of skill and speed. The NEIC gets data from nearly 1,800 stations worldwide, but there are gaps that slow the seismic analyses. China's national seismological alerting network puts a 30-minute delay on much of the information, so Earle's team can rarely use it. And India does not release its seismic data. Nepal, where seismologists have long warned about the earthquake risk, did not have a single station feeding real-time data into the USGS system. Had the agency received more real-time data from locations closer to the epicentre, seismologists could have accurately located the Nepal quake faster than they did, says Thorne Lay, a seismologist at the University of California, Santa Cruz.

Even with all its speed, the NEIC is not the fastest earthquake-alert system in the United States. That title goes to the National Oceanic and Atmospheric Administration's two tsunami-warning centres. Drawing on the same seismic network, they release rougher magnitudes and locations within 3 minutes of an earthquake striking, but they handle only shocks in oceans near US territory.

The NEIC keeps pushing to shave as many seconds off its own notifications as possible. One ongoing project involves Twitter. Earle has set up an automated system that hunts for words such as 'earthquake' in various languages in tweets from around the world (P. Earle *Nature Geosci.* **3,** 221–222; 2010). He has to filter out unrelated instances, including references to the video game *Quake*, but once that is done he can get a heads-up that something big is beginning. When someone in Indonesia tweets '*gempa*', or earthquake, "it's on our server in five seconds," he says.

Tweets can arrive at the NEIC faster than seismic waves can reach recording stations. In 2012, a magnitude-4.0 jolt in Maine set off a stream of tweets from the region around the epicentre. Earle got an automatic text notification before the shaking spread across New England. "I was at Safeway buying groceries, and I knew about the quake, from nothing but Twitter data, before other people felt it," he says.

The Twitter experiment is most useful in places where the USGS does not receive a lot of real-time data, such as parts of South America or Indonesia. Although it will never replace the NEIC's conventional methods, it can alert the seismologists there to keep a lookout for incoming data.

The earthquakes never stop coming. Towards the end of a long Friday afternoon in May, Earle is at his standing desk when his iPhone buzzes with a report of a magnitude-6.9 quake in the Solomon Islands. "That one isn't going to be near a populated area, but it's a big quake," he says. "I'm gonna get someone." He is heading out of the door nearly before he finishes the sentence.

Earle speed-walks down the hallway, past the row of display monitors set up for television cameras, and pokes his head into the office of seismologist Jana Pursley. "Jana, have you got that?" he asks. "No, Sean does," she says, waving her hand at the on-duty seismologist down the hall. "OK," says Earle. "Sean will release it, and then I'll have Bruce review the moment tensors for it, and then we'll be done."

With that earthquake sorted, Earle heads back to his office. He switches on the electric kettle that sits next to two containers of freeze-dried, generic-brand coffee. "I get the cheapest possible coffee because I don't even taste it anymore," he says. "I just drink it."

And he turns back to his monitor, to wait for the next one. ∎

Other groups also rolled into action. Gisli Olafsson in Reykjavik, who directs emergency response for a consortium of 43 humanitarian groups called NetHope, says: "I always look at PAGER once it becomes available." Studying the USGS information, he was relieved to see that the shock had originated relatively far from Kathmandu. But he also learned that the quake had struck in mountainous terrain on a fault close to Earth's surface, which meant that it had probably destroyed roads. NetHope immediately started preparing for the complicated logistics of getting in and out of rural areas with limited access, and Olafsson flew to Kathmandu to coordinate its response.

Even the financial world got involved: the Inter-American Development Bank uses PAGER numbers to trigger payouts on catastrophe bonds, a type of insurance against natural disasters such as earthquakes.

The most recent estimates suggest that the 25 April earthquake and its aftershocks, including a magnitude 7.3 on 12 May, killed roughly 8,700 people — close to the PAGER estimates of around 10,000 deaths. Other catastrophe experts had estimated 50,000 dead or more, using independent assessments of population exposure and building vulnerability.

One factor that may have saved lives in Kathmandu was how buildings were constructed, says Kishor Jaiswal, a civil engineer at the NEIC. Many of the newer buildings in the city have concrete frames reinforced with steel bars, which kept a lot of them from collapsing. Jaiswal had previously analysed this construction, and his work was one reason that the PAGER fatality estimates were relatively low. Although the toll was great, he knew that much of the city would survive.

NEED FOR SPEED

Most of the NEIC's work is much calmer than on the night of the Nepalese disaster. Of the thousands of earthquakes that the team tracks every month, the vast majority do not kill anyone. Earle, Benz and Hayes spend their time developing ways to analyse earthquake ruptures as quickly and accurately as possible. Hayes, for instance, specializes in 'moment tensor' and 'finite fault' calculations, both of which convey information about exactly how a fault has ruptured.

One of Earle's top priorities for the earthquake centre is to avoid making major mistakes, although his team sometimes does err. Notable bloopers include issuing an alert on Christmas Day 2013 for a magnitude-22 earthquake. It was supposed to say magnitude 2.2; the typo caused the NEIC to remove all human typing from the real-time system.

And in May this year, the USGS reported several phantom quakes in California — in reality, they were vibrations from more-distant shocks

Alexandra Witze *writes for* Nature *from Boulder, Colorado.*

BARRY GUTIERREZ

How to beat HIV

Scientists have the tools to end the epidemic. They just need better ways to use them.

BY ERIKA CHECK HAYDEN

On the shores of Lake Victoria, Kenyan fishermen spread out their nets on the sand to dry their catch in the sun. At a clutch of tents next to the beach, health-care workers are casting a very different kind of net, one that could help to capture the best approach to eradicating HIV.

The tents draw a steady stream of visitors because the fishermen and their families, as well as farmers, students and others from the surrounding communities, have heard that they can get vitamin A, condoms, and medicines for worms and malaria there. At the same time, they are offered various screening tests — including one for HIV. The hope is that, along with taking advantage of the other medical services, they will agree to be tested and, if necessary, treated for the sexually transmitted virus.

Here in Kenya's Nyanza Province, which has the country's highest rate of HIV infection, this community is part of a groundbreaking study designed to explain a troubling conundrum. Interventions to prevent HIV transmission that work in trial settings — such as rapid on-the-spot HIV tests coupled with effective treatments — often fail to make as much of a dent in the epidemic as they should. The current trial, known as Sustainable East Africa Research in Community Health (SEARCH), has enrolled more than 335,000 people in Kenya and Uganda and is at the forefront of a shift in thinking about how best to deal with HIV. In the past, there was a sense that stopping the HIV/AIDS epidemic would require some radically new biomedical intervention, such as a cure or a vaccine. The growing

consensus, however, is that the tools needed to stamp out HIV already exist if they could just be used in the right way.

In trials over the past decade, experimental interventions such as voluntary male circumcision or the use of prophylactic drugs produced head-turning results that earned them funding for broader implementation. But they have not succeeded when rolled out more generally: in some cases because the funding did not last, but in others because the conditions of a clinical trial are not the same as those in real life. SEARCH and efforts like it are intended to explain why. They fall within the domain of implementation science, an emerging multidisciplinary field that seeks to understand and overcome factors — such as human behaviour and economics — that can lessen the impact of interventions that have otherwise proved effective.

Major aid programmes are taking an interest. The US President's Emergency Fund for AIDS Relief (PEPFAR), for example, launched a US$60-million programme in implementation science in 2012. Among other aims, this programme is testing whether integrating the prevention and treatment of HIV infection with other facets of countries' health and social systems — such as family planning, tuberculosis treatment and education — could help to get the HIV epidemic under control.

"A lot of my university colleagues are very good at doing the studies and coming up with a finding, but are clueless about how to get that finding into actual practice," says epidemiologist Farley Cleghorn of the Futures Group in Washington DC, which contracts with

governments to conduct aid programmes. "The challenge for implementation science is to diminish that reduction in impact that happens when you move from a controlled environment to the general population."

AMBITIOUS GOALS

SEARCH fits into a bold global strategy for eradicating HIV. In 2014, the Joint United Nations Programme on HIV/AIDS (UNAIDS), based in Geneva, Switzerland, laid out the '90-90-90' target: getting a diagnosis for 90% of people infected with HIV; putting 90% of those on antiretroviral therapy; and getting 90% of those virally suppressed, meaning that they have an undetectable level of HIV in their bodies. Achieving these goals by 2020 would herald an end to the epidemic as a global threat by 2030, with the number of new infections per year limited to about 200,000.

That is easier said than done, however. "To say it's an ambitious target would be an understatement," says Mitchell Warren, director of AVAC, an AIDS prevention advocacy group. Less than half of the people with HIV in some areas of the world, such as southern Africa, have access to HIV tests. In most regions, less than 40% of people with HIV are being treated, and the percentage of people with HIV who are virally suppressed is quite low in many regions (only about 30% in the United States, for example). Worldwide, about 15 million people will have access to antiretroviral treatment in 2015 (see 'Signs of change').

The problem is that individuals drop out at each step of the path that leads to viral

suppression. Most people with HIV have never been tested. Of those who have, many do not start treatment; and of those who do, many stop for a variety of reasons. Implementation science is finding that some of the best ways to plug the holes in this leaky cascade of care are to make it easier and more rewarding for patients to get the medical attention they need.

The problem is acute in sub-Saharan Africa, which represents 70% of the global total of both new infections and people living with HIV. Eight years ago, doctors with the aid group Médecins sans Frontières (MSF) noticed that, as the number of people being treated with antiretrovirals increased, patients attending an HIV clinic in the township of Khayelitsha in Cape Town, South Africa, were finding it increasingly difficult to get their medication. To pick up their pills, they had to visit the clinic for frequent check-ups and tests of their viral load and T-cell count — indications of the progression of the infection. But at every appointment, they faced hours-long waits to see overburdened nurses. And people frequently left empty-handed owing to a shortage of the drugs. As many as one-quarter of patients who started HIV treatment stopped after one year.

MSF decided to try something different: it set up clubs that met every two months at the clinic, led by trained counsellors, many of whom were patients themselves. The clubs met during slow times at the clinic, and counsellors brought each patient's supply of medicines to the meeting in a pre-packed bag and led a group discussion about the importance of staying on treatment. A nurse visited once a year to take blood samples and measure viral load and T-cell count.

The clubs were a dramatic success: for the patients who received their care in this way, there was a 57% decrease in the number of people dropping out (through either death or giving up treatment) compared with the group that continued to receive care through the previous system at the clinic itself[1]. Such clubs are now seen as a model of how to keep patients in care and have been organized in less formal settings such as private homes after work hours.

The SEARCH trial is building on the idea of adapting the care of people with HIV to their needs. It is taking a broader look at the problem by not only bringing care closer to patients and making it easier for them to get it, but also examining whether integrating HIV care into the overall health-care system can help stop the leaks at each step of the care cascade.

The first step is diagnosis. As few as 40% of Kenyans infected with HIV know that they have it. One problem has been that people eschew targeted HIV-testing campaigns. Another is that those most likely to be infected, such as people who migrate to find work, are least likely to be reached by testing campaigns. So SEARCH is evaluating other ways to attract people — for example, by deploying community health campaigns such as the one in Nyanza Province, where people can access much-desired medical services as well as HIV tests. People who do not attend the community programmes are approached through door-to-door campaigns and are offered HIV tests that they can take in their own homes.

This combination of mobile campaigns followed by home visits has boosted the proportion of adults who had taken at least one HIV test from 57% to 80% in the communities included in SEARCH, Gabriel Chamie, at the University of California, San Francisco (UCSF), reported on behalf of the study at a conference in February.

As the SEARCH trial progresses, it will assess ways to get those who test positive more quickly into care and to keep them there. They are started on antiretroviral drugs rapidly — sometimes on the same day as their diagnosis. The project has enacted a triage system for speeding HIV patients who feel well into and out of the clinic when they attend appointments, and for reducing the number of visits. The project is also trialling appointment reminders and is setting up a telephone hotline to help keep patients engaged in their care. And the project will measure a person's viral load at the start of treatment, six months later, and each year subsequently, to check whether the treatment is working.

"What we've tried to go do is greatly simplify HIV care delivery," says Diane Havlir at UCSF, one of the directors of the SEARCH study.

Researchers are also using implementation science to understand why prevention methods such as circumcision and prophylactic drug treatment have not been adopted as widely as they could have been.

For instance, trials in the mid-2000s proved

Informal community groups deliver HIV therapy in South Africa.

that voluntary circumcision for men cut the risk of their acquiring HIV from a female sexual partner by 60%. The World Health Organization recommended in 2007 that circumcision be used for prevention and, with UNAIDS and the Bill & Melinda Gates Foundation in Seattle, Washington, set a target to circumcise 80% of eligible men in Africa by 2016 to prevent up to 3.4 million new HIV infections (see *Nature* **503**, 182–185; 2013). PEPFAR and others provided funding, and 9 million circumcisions have been performed since 2007.

But even this massive campaign has up to now reached only 28% of its target. One problem is that circumcision is a surgical procedure and so requires different expertise and resources from those in current HIV programmes. And setting up stand-alone circumcision programmes diverts resources from existing surgery, which is already under-resourced. "There's a whole lot of logistical and operational issues that are resulting in countries not meeting their targets," Cleghorn says.

A different set of real-world issues has complicated what is known as pre-exposure prophylaxis (PReP) — the concept of taking a dose of antiretroviral medication regularly or around the time of sexual intercourse to prevent infection. In the PROUD study in the United Kingdom, which reported results in February, this has been shown to reduce the risk of infection by 86% in men who have sex with men, and studies of PReP in Africa showed a decrease of 73% in heterosexual couples[2].

But despite these results, PReP is not widely used. One reason is that some people at highest risk of becoming infected with HIV are also those most likely to be in denial about their risk, or unable to access services, and are therefore least likely to take a medicine to prevent infection. And developing countries have enough difficulty distributing antiretrovirals to people already infected to make a serious effort to give them to anyone else. In June, for instance, MSF reported that one in three health facilities in South Africa reported a shortage of medications for HIV or tuberculosis late last year. That makes it hard to face the additional challenge of getting drugs to those who are HIV negative. "They can't wrap their heads around it," Cleghorn says.

In addition, PReP has consistently failed to protect those arguably most in need of new prevention options — young unmarried women. In most of the poor countries hit hard by HIV, 80% of new infections among adolescents are in girls. Yet PReP has failed in this demographic in trials that used many different delivery approaches, such as vaginal gels containing antiretroviral medication or oral pills taken daily or before and after sex.

The main problem is that many women did not use the products they were given. In one study of 5,000 women in South Africa, Zimbabwe and Uganda, blood tests showed that

SIGNS OF CHANGE

The past decade has seen some success against HIV, with the number of new infections and deaths in decline. The progress has inspired more-ambitious targets for the next five years that would herald the end of HIV as a worldwide threat by 2030.
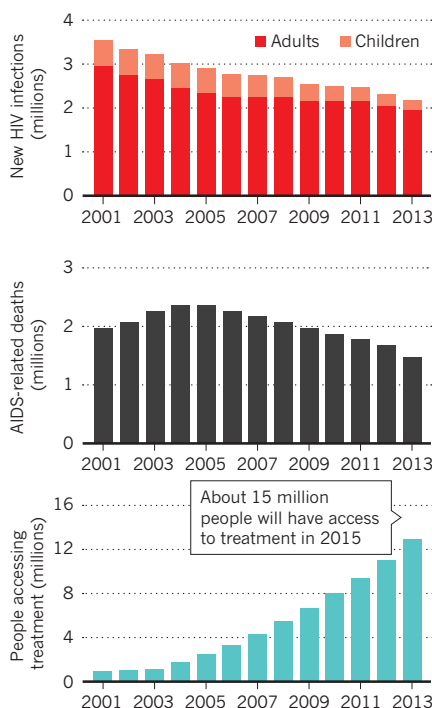
only 25–30% of participants actually used the medications, even though 88% said that they had. Those questioned in small groups said that they did not use the products because of social factors, such as fear that they would be ostracized or perceived as having HIV already if they were known to possess HIV drugs.

The problem is part of a broader social context that makes girls more vulnerable to HIV than boys of their age. Many date older men, who have a higher prevalence of HIV infection than adolescent boys; some engage in transactional sex to afford necessities; and some are abused.

IMPLEMENTING SOLUTIONS

Implementation science is trying to find ways to address these broader factors in an attempt to cut the HIV risk in girls. In a meta-analysis published in March, social scientist Nicole Haberland of the Population Council in New York City examined programmes designed to reduce pregnancy, HIV and sexually transmitted disease infection rates in girls[3]. She found that when these programmes included educational components that specifically addressed gender or power — for instance, by including discussions of how girls could negotiate condom use and how gender inequality influenced their own lives — they were more likely to reduce disease risk. Eight of 10 programmes that included such components worked, compared with 2 of 12 that did not address these issues.

Responding to findings such as these, in December 2014, PEPFAR announced the DREAMS initiative which, in conjunction with the Bill & Melinda Gates Foundation and the Nike Foundation, will spend $210 million over two years to provide a combination of preventive interventions targeting young girls, such as HIV testing, counselling and care for rape survivors, and programmes aimed to boost the resilience of girls and their families, such as cash payments for girls who stay in school.

But drawing a direct link between some of these interventions and lowering the risk of HIV infection in girls has been difficult. Two studies that are specifically testing whether cash transfers for children who meet certain academic goals can cut the risk of new HIV infections in South Africa are expected to report their results at the upcoming meeting of the International AIDS Society in Vancouver, Canada, on 19–22 July.

Epidemiologist Audrey Pettifor, who leads one of the trials, says that although such interventions have worked in very poor countries — such as Malawi — they may not apply elsewhere. In her trial, girls and their families were paid the equivalent of $24 per month if the girls attended school, but the youngsters in South Africa have very different expectations from those in much poorer African nations. Along with high levels of poverty, unemployment and HIV prevalence is a desire for luxury goods — the girls in Pettifor's trial named items such as designer jeans, Italian shoes and Blackberry smartphones as necessities. "If we're trying to deter transactional sex, it's going to be a big ask," Pettifor says. It may not work.

Implementation science is still relatively new to the HIV/AIDS field, and it is not yet clear if it will help researchers to hit all of the 90-90-90 goals. "The evidence base is still mixed on programmes or interventions to reach these goals," Pettifor says.

Researchers hope that the field will mature and become more rigorous. The SEARCH trial, for example, is assessing whether streamlining HIV care has knock-on health and economic benefits for the community, such as elevated fishing or farming revenues, or enhanced education rates among children.

The fish catch of a small community on the shores of Lake Victoria may seem far removed from the goal of stopping HIV — but implementation scientists see it as an essential part of the work. "We've set these very aspirational goals," says Havlir. But if they want to reach them, then scientists must get to grips with the complexities of the real world. ■

Erika Check Hayden writes for Nature *from San Francisco.*

1. Luque-Fernandez, M. A. *et al. PLoS ONE* **8,** e56088 (2013).
2. Baeten, J. M. *et al. N. Engl. J. Med.* **367,** 399–410 (2012).
3. Haberland, N. A. *Int. Perspect. Sex. Reprod. Health* **41,** 31–42 (2015).
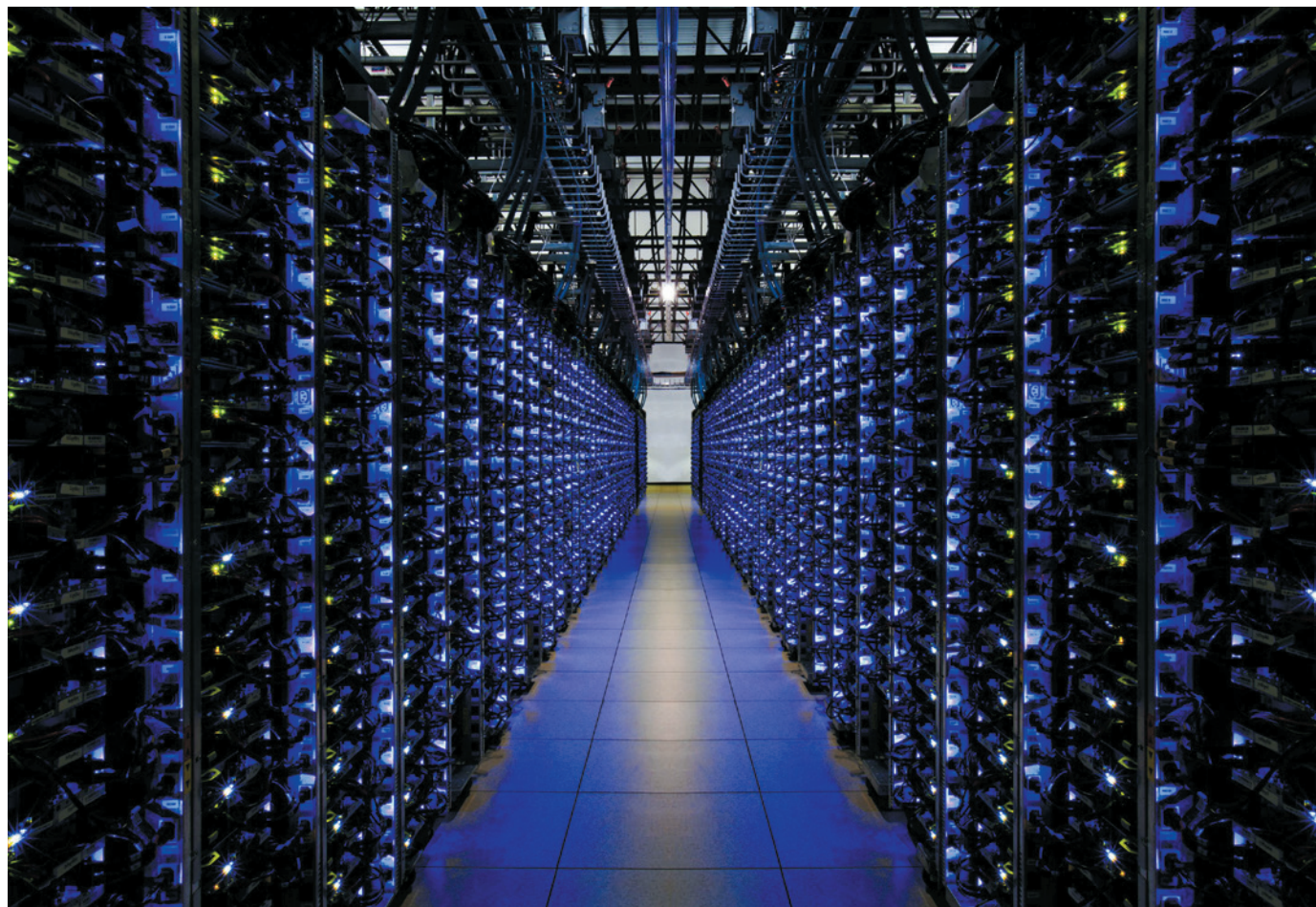
COMMENT

Google's cloud services are among those increasingly being used by researchers who want to analyse large genomics data sets.

Create a cloud commons

Major funding agencies should ensure that large biological data sets are stored in cloud services to enable easy access and fast analysis, say **Lincoln D. Stein** and colleagues.

There was a collective cheer in the human genomics community earlier this year, as researchers — ever more stymied by the challenges of accessing vast data sets — saw a major roadblock disappear. In March, the US National Institutes of Health (NIH) lifted its 2007 restriction on the use of cloud computing to store and analyse the tens of thousands of genomes and other genetic information held in its repository, the database of Genotypes and Phenotypes (dbGaP)[1].

Cloud services offer customers large amounts of storage and computing power on a pay-as-you-go basis. Because these services are available through the Internet, and multiple users share hardware, numerous funding agencies have been concerned that their use in genomics could threaten the privacy of people who supply samples[2].

The NIH turnaround is part of a growing suite of efforts aimed at addressing the fact that in the human genomics research community, the challenges of accessing big data sets are now blocking scientists' ability to do research, and especially to replicate and build on previous work (see go.nature.com/h9jgs1).

To take full advantage of the possibilities that cloud computing offers, we ▶

▶ urge the NIH and other agencies to pay for the storage of major genomic data sets in the most popular cloud services. This way, instead of thousands of researchers wasting time and money by independently transferring data from a repository to the cloud of their choice, authorized scientists would be able to tap easily and cheaply into a global commons as and when they need to.

BIG DATA

Thanks to improvements in sequencing technology, the volume of genomic data submitted to public archives is now well into the multi-petabyte range (1 petabyte is 10^{15} bytes). In the International Cancer Genome Consortium (ICGC)[3], for instance, groups from 17 countries have amassed a data set in excess of two petabytes — roughly 500,000 DVDs-worth — in just five years.

Using a typical university Internet connection, it would take more than 15 months to transfer a data set this size from its repository into a researcher's local network of connected computers. And the hardware needed to store, let alone process the data, would cost around US$1 million.

Cloud services provide 'elasticity', meaning that a researcher can use as many computers as needed to complete an analysis quickly, and pay for only the computing time used. Several researchers can work in parallel, sharing their data and methods with ease by performing their analyses within cloud-based virtual computers that they control from their desktops. Thus the analysis of a big genome data set that might have previously taken months can be executed in days in or weeks (see 'Express lane').

These days, cloud services are just as secure as most academic data centres, often more so. They are now offered by major commercial companies including Amazon, Google and Microsoft, as well as smaller companies focused on genomics research, such as California-based Annai Systems, and several academic institutions such as the European Bioinformatics Institute in Hinxton, UK. These providers use strong encryption for data, have systems, such as firewalls and keychain fobs, for controlling who has access to the data, and provide tools to the owners of the data that allow them to monitor use closely.

A few major funders of human-genomics research are being cautious — for instance, some European funding agencies recommend that researchers keep genomic data within the agencies' jurisdiction to comply with European law on privacy[4]. But the cheapness, flexibility, reliability and security of cloud computing is such that we anticipate a wholesale shift to cloud services over the coming months (see 'Reaching for the cloud'). And we welcome the NIH's decision in hastening this transition.

Now is the time to establish mechanisms and practices that maximize the efficiency and usability of cloud computing while minimizing costs.

> *"The human genomics community could pave the way for other researchers grappling with data overload."*

ACCESS CONTROL

To gain access to much of the human genomic and other data held in central repositories such as the dbGaP or its counterpart, the European Genome-phenome Archive (EGA), a researcher must obtain approval from a data-access committee, or DAC. Currently, if two independent research groups wish to work on the same data set in a private or commercial cloud, they will each need to get approval from the relevant DAC, copy the data across the Internet and store it in their cloud of choice.

Both groups have to wait while the data are copied, and each has to pay for storage while the data are being copied and for as long as they need the data. As hundreds of groups start to do the same thing, this process could collectively waste years of researchers' time and tens of millions of taxpayer dollars. Even with unfettered access to cloud services, it is currently impractical for most groups to work with the largest public genomic data sets because of the time and costs involved in transferring the data from its repository into a cloud.

A better approach would be for the relevant funding agencies to request that every major genomic data set be uploaded into the most popular academic and commercial clouds available, and to pay for the long-term storage of the data in the clouds. This way, the data would need to be copied only once and researchers would have to pay only for the temporary storage they use while their analysis is in progress.

Currently, several commercial providers of cloud services are offering to store research data sets for free or at heavily subsidized rates to prompt more researchers to use their services. Amazon Web Services, for example, levies no charge for hosting the sequences released by the 1,000 Genomes Project (now totalling more than 200 terabytes of data), an international effort to catalogue human genetic variation. And Annai Systems hosts a growing subset of the ICGC data set.

We envisage that entities such as the dbGaP or the EGA[5] would continue to be the primary custodians of the data and that their DACs would still review and authorize data use within the cloud. In this way, genomic cloud computing could even give rise to a micro-economy. For instance, a genome biologist who contributes a valuable data set to a cloud could receive credits for processing time. Similarly, a computer scientist who contributes a software package that enables other geneticists to find cancer variants more efficiently, say, could receive credits every time someone runs their package.

Over time, a virtuous cycle would result. Being able to merge large data sets would enable researchers to link rare genetic variations to diseases, and such successes would encourage others to deposit more data sets and the development of yet more powerful software. Such mechanisms could work in conjunction with requests from funding agencies that certain data sets be deposited in certain clouds.

One possible risk is that, by rising to dominance, a single provider of cloud services could come to control pricing,

REACHING FOR THE CLOUD

Internet cloud services, which provide large amounts of data storage and computing power, are becoming increasingly popular with geneticists grappling with vast data sets.
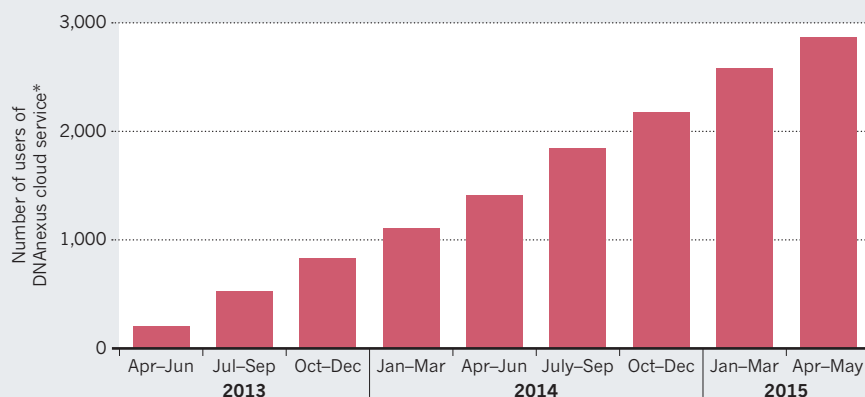
*Data from DNAnexus, a cloud-based genome informatics and data-management platform.

SOURCE: DNANEXUS

EXPRESS LANE

The Pan Cancer Analysis of Whole Genomes project (in which L.D.S., P.C., G.G. and J.O.K. are involved), an effort to investigate the role of non-coding parts of the genome in cancer, demonstrates how much faster and cheaper it is to use cloud computing than to use conventional academic data centres when analysing vast biological data sets.

2,617 PATIENTS

Researchers are using cloud computing to analyse 500 patient samples, while academic data centres are being used to analyse 2,117 samples, owing to funding-agency restrictions on the use of cloud services.

600+ TERABYTES
(1 terabyte = 10^{12} bytes)

CLOUD COMPUTING

500
Samples analysed so far

US$18
Cost of analysing one sample

333
Samples analysed per month

6 WEEKS
Time taken to complete analysis

For each patient, the genomes of a tumour cell and a normal blood cell are compared.

10 ACADEMIC DATA CENTRES

1,827
Samples analysed so far

US$200
Cost of analysing one sample*

30
Samples analysed per month per centre

6+ MONTHS
Time taken to complete analysis

*If using a standard university computer system and buying the hardware.

---

will also be needed. Individual project DACs should continue to be gatekeepers in the short term, but ultimately a few 'general-purpose' DACs may be better placed to oversee access to the clouds than the multitude of DACs currently operating.

On the legal side, rules of the road must be established to clarify the roles and responsibilities of the funding agencies, the data custodians, the cloud service providers and the researchers who use cloud-based genomic data. If someone posted an ICGC genome on Facebook, for instance, who among these various players should be held accountable? Fortunately, for the past two years, an international coalition, the Global Alliance for Genomics and Health (genomicsandhealth.org) has prepared a Framework for Responsible Sharing of Genomic and Health-Related Data.

Meanwhile, the US National Cancer Institute has several pilot projects[3] exploring the practicalities of sharing and analysing genomic data on clouds. And the NIH and other funding agencies are already discussing a variety of 'biomedical commons' concepts, which incorporate several of the ideas proposed here.

By taking the right approach to cloud computing, the human genomics community could pave the way for researchers in many other fields, from neuroscience to epidemiology, who are similarly grappling with data overload. ∎

Lincoln D. Stein *is director of informatics and biocomputing at the Ontario Institute of Cancer Research, Toronto, Canada.* **Bartha M. Knoppers** *is director of the Centre of Genomics and Policy, McGill University, Montreal, Canada.* **Peter Campbell** *is head of cancer genomics at the Wellcome Trust Sanger Institute, Hinxton, UK.* **Gad Getz** *is director of the Cancer Genome Computational Analysis group at the Broad Institute of MIT and Harvard, Cambridge, Massachusetts, and is the director of bioinformatics in the Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA.* **Jan O. Korbel** *is a group leader in genome biology at the European Molecular Biology Laboratory, Heidelberg, Germany.*
e-mail: lincoln.stein@gmail.com

1. Tryka, K. A. *et al. Nucleic Acid. Res.* **42,** D975– D979 (2014).
2. Gutmann, A. & Wagner, J. W. *Hastings Cent. Rep.* **43,** 15–18 (2013).
3. The International Cancer Genome Consortium. *Nature* **464,** 993–998 (2010).
4. European Commission. *Eleventh Annual Report of the Article 29 Working Party on Data Protection* (European Commission, 2008).
5. Leinonen R. *et al. Nucleic Acid. Res.* **39,** D28–D31 (2011).
6. Brazma, A. *et al. Nature Genet.* **29,** 365–371 (2001).

---

and so subtly influence how the science is performed. To prevent this from happening, funding agencies should fund the deposition of the same important data sets in multiple clouds. This would also help to address jurisdictional sticking points. Genomic data originating in Europe, for instance, could be confined to clouds based in Europe.

GENOMIC STANDARDS

Achieving this vision will require work, technical and legal. For example, currently there is no way for a cystic-fibrosis researcher, say, to write software to search the dbGaP database and find the sequences obtained from people with the disease. Systematically tagging the data — specifying the tissue source of the sample, for instance — would help to address this. Since 2001, journal publishers have agreed to accept only RNA microarray studies in which researchers describe their data using the 'minimum information about a microarray experiment (MIAME)' standard[6]. A similar standard is needed for genomic data.

Reliable protocols for authorizing access to sensitive data in the cloud, as well as mechanisms to enable and revoke access

A farmer in Burkina Faso improved his livelihood by using a water pump to irrigate his land.

Development goals should enable decision–making

Gathering data that answer particular questions is the most effective way to support the Sustainable Development Goals, say **Keith Shepherd** and colleagues.

In September, a United Nations summit of heads of state will adopt the Sustainable Development Goals (SDGs) — a set of 17 goals and 169 targets to guide international development. A diverse range of indicators and monitoring strategies is being proposed, covering every dimension of development, from human well-being to the environment[1].

Next week, high-level political representatives meeting in Addis Ababa for the International Conference on Financing for Development will discuss how to fund the SDGs. The participating governments, development institutions, non-governmental organizations (NGOs) and business stakeholders will negotiate an agreement on domestic commitments and international action around financing initiatives.

The SDG monitoring framework makes great demands on nations — it must help countries to implement strategies and allocate resources, measure progress towards sustainability and hold stakeholders to account[1]. A country found to be failing in sustainable forestry, for example, may choose to invest more in forestry or receive penalties and lose aid. Target-setting is trendy among aid and development organizations as well as in multilateral agreements for accountability, impact and value for money.

We contend that target-setting is flawed, costly and could have little — or even negative — impact.

First, targets may have unintended consequences. For example, education quality as a whole suffered in some countries that diverted resources to early schooling

to meet the target of Millennium Development Goal (MDG) of achieving universal primary education[2].

Second, target-setting inhibits learning by focusing efforts on meeting the target rather than solving the problem[3]. The milestones are easily manipulated — aims such as halving deaths from road-traffic accidents can trigger misreporting if the performance falls short or encourage underperformance if the goal can be exceeded.

Third, it is costly: development partners will have to reallocate scant resources for a 'data revolution' that will cost an estimated US$1 billion a year[4].

We advocate a different approach. Governments and the development community need to embrace decision-analysis concepts and tools that have

been used for decades in mining, oil, cybersecurity, insurance, environmental policy and drug development[5,6]. Our call to adopt this approach is based on five principles.

FIVE PRINCIPLES

Replace targets with measures of investment return. The SDGs should state a few broad strategic goals and assess how to achieve them by measuring each project in terms of a return on investment: how well the goals are met given the resources used. For example, were the environmental benefits and reduction of poverty enough to justify the allocation of limited resources?

Decision-makers would use economic models that project long-term costs, benefits and risks of intervention options. They would seek to maximize the risk–return position of a portfolio of options towards achieving the development objectives[5]. This will require the relative value of different aims to be stated in monetary terms. A government could assess, for instance, whether its objective would best be achieved by spending $50 million on training farmers, building roads, improving education or some combination of them.

Model intervention decisions. Enabling decision-making must be at the heart of SDG monitoring strategies. It is difficult, however, to pinpoint which data are required to support better decision-making without formal decision analysis.

For example, public-health scoring systems — such as the Framingham Risk Score for cardiovascular disease — that assess and prioritize patients according to factors such as age, blood pressure and cholesterol level do not account for people with the most susceptibility who have received treatment. The scoring system underestimates the risk factors if treatment is not recorded, no matter how many other data are collected[6].

In 2013, we conducted a survey[7] of 110 stakeholders in African agriculture (including scientists, universities, donors, government ministries, NGOs, the private sector and farmer associations). Most (54%) could not identify a policy or management decision that would be supported by further data. They might say, for example, that better soil data would help them to manage erosion-control policies better, but they could not name a particular decision, investment, intervention or policy that would be different if they knew more about the soil. Only 15% of respondents were able to articulate how acquiring data would reduce a crucial uncertainty to enable a decision.

The survey showed that there was a tendency, especially among scientists, to seek data for the sake of having them. For example, biodiversity and poverty data were frequently cited as a focus of effort but infrequently as a

Water-pipeline planning could be improved by incorporating decision-focused data.

perceived need or uncertainty. Climate data were needed and satisfied an uncertainty, but were infrequently collected.

The SDG community must define the actions, policies, programmes or projects that the indicators are expected to inform. These should reflect the practical choices that development planners on the ground will face, such as whether to build one large dam or many small ones to secure water and energy needs, or which of several child-nutrition programmes should be implemented in a region.

The impact of interventions on different groups of people should be factored in: for example, upstream and downstream water users, male and female farmers, or rural and urban populations may be affected differently by a given policy. Such a user-centred approach[8] to deciding the best actions would make decision-makers' assumptions and preferences transparent — for instance, the degree of risk they are willing to accept.

"Decision-makers who are implementing and tracking the SDGs should employ probabilistic decision analysis."

Integrate expert knowledge. It is a common mistake to assume that 'evidence' is the same as 'data' or that 'subjective' means 'uninformative'. Decision-making should draw on all appropriate sources of evidence. In developing countries where data are sparse, expert knowledge can fill the gaps. For instance, in our assessment of the

viability of agroforestry projects in Africa, we used our experience to set ranges on tree-survival rates, costs of raising tree seedlings and farm prices of tree products. Decision theorists and local experts will have to work together to identify relevant variables, causal associations and uncertainties.

There are well-established procedures for 'calibrating' experts when using subjective probabilities to quantify uncertainty about estimates[5,6]. For example, the World Agroforestry Centre assessed the relative benefits of agricultural interventions for developing regions by calibrating experts for how well they estimated probabilities and by holding workshops to define a probabilistic model[5].

The most widely accepted method of incorporating knowledge for probability assessment is Bayes' theorem. This updates the likelihood of a belief in some event (such as whether an intervention will reduce poverty) when observing new evidence about the event (such as the occurrence of drought)[6]. Bayesian analyses — incorporating historical data and expert judgement — are used in transport and systems-safety assessments, medical diagnosis, operational risk assessment in finance and in forensics[6], but seldom in development. They should be used, for example, to evaluate the relative risks of competing development interventions.

Include uncertainty in predictive models. Scientists often use simulations of climate, hydrology, crop growth or disease spread to guide policy or management decisions. Such models of physical systems have two limitations for allocating resources. First, they usually omit behavioural and economic ▶

▶ factors; and second, they commonly fail to represent uncertainty in input data, model parameters and outputs.

Decision-makers who are implementing and tracking the SDGs should employ probabilistic decision analysis, for example Monte Carlo simulations[5] or Bayesian network models[6]. Provided that such models are developed using properly calibrated expert judgement and decision-focused data, they can incorporate the key factors and outcomes and the causal relationships between them. For instance, simulations for evaluating options for building a water pipeline could take into account rare 'what-if' scenarios, such as a hurricane during development, and predict (with probabilities) the time and cost of implementation and the benefits of improved water supply.

Measure the most informative variables.
An analysis of more than 80 models from a variety of decisions and industries reveals that managers tend to choose to measure variables that are unlikely to improve decisions while ignoring more useful ones[5]. For example, the adoption rate of a method by farmers is easy to measure, but its effect on yields may be more relevant for making choices. Quantities for which there is already a great deal of information, such as financial costs, are more likely to be tracked but cannot influence decisions because there is little left to learn about them. Less common variables such as social and long-term benefits (such as on mental health) and environmental impacts (such as water pollution from soil erosion) may be of greater value.

Reducing decision uncertainty should be the purpose of measurement[5]. Only a few variables may be relevant, and data collection should focus on those that narrow choices the most[5]. For example, a US Environmental Protection Agency analysis of alternative information systems for water quality found that only one variable dominated the uncertainty around investment in the information system: the average health effects of safe-drinking-water policies. Uncertainties about adoption rates of the technology, efficiency improvements and improved reporting rates turned out to have no information value for the agency[5].

In decision theory, the value of information is the amount that a rational decision-maker would be willing to pay for that knowledge before making a decision — the value of clairvoyance[9]. This can be estimated only by analysing the uncertainties in all the variables that have a bearing on a decision. Such value-of-information analysis is not used in development but is in, say, health economics[10]. The UK National Institute for Health and Care Excellence uses it in deciding whether a drug or intervention should be approved for widespread use[10].

Some proposed SDG indicators will be difficult and expensive for low-income countries to collect, for example the "percentage of women, men, indigenous peoples, and local communities with secure rights to land, property, and natural resources", and "nitrogen use efficiency in food systems". Limited resources would be better spent on gathering data with high decision-making value. Those data can be identified only by analysing the specific decisions to be made, and will change as new decision nodes emerge.

Value-of-information analysis helps to identify metrics for monitoring performance. These are often not intuitive and therefore missed. For example, we did a study of natural-resource management interventions, such as integrated watershed projects and seed improvements for maintaining agro-biodiversity. We found that the most useful factors to know were rural-to-urban migration rates, market prices, project failure risks, negative consequences (such as disadvantaging poorer sectors of the community) and adoption rates[5].

A NEW DIRECTION

Decision analysts should be embedded in all government and UN policy-development and management units, through a capacity-development programme paid for by governments and international donors, including from the private sector. The UN should establish a forum of decision-analysis experts to steer this initiative.

"Reducing decision uncertainty should be the purpose of measurement."

These analysts would work with decision-makers and subject experts to clarify key intervention decisions and develop probabilistic models of alternative actions. They would build models in a participatory way, involving key stakeholder groups and training experts in subjective probability estimation.

Value-of-information analysis should guide data-collection efforts and define high-value metrics that have the potential to improve decisions and performance. Some of the proposed SDG indicators might be among them, but would be rationally justified, and may change as new priorities emerge.

For commonly occurring variables, such as carbon and commodity prices and risks of extreme climate events, governments and the UN should establish open-access libraries of probability distributions for running simulations[5]. Monitoring real change against decision models provides a realistic alternative in circumstances in which it is difficult to conduct randomized control trials, such as when considering major new environmental interventions.

We call on the delegates of the Financing for Development conference in Addis Ababa to establish a task force to explore our approach. We recommend that some of the aid money earmarked for improved monitoring of the SDGs be directed to establishing this initiative. Forward-looking governments, especially in data-sparse countries, should consider pioneering decision-analysis approaches.

The principles that we have outlined are applicable to the improvement of any policy or management process, from international policy (such as climate-change negotiations) down to the individual project level (such as whether a village should install a new water storage system). Training a generation of decision analysts to work with policy-makers could do more for development than any other single intervention. ■

Keith Shepherd *is leader of the Science Domain on Land Health Decisions at the World Agroforestry Centre, Nairobi, Kenya, and co-leader of Decision Analysis and Information Systems in the Research Programme on Water, Land & Ecosystems of the Consultative Group for International Agricultural Research, Montpellier, France.* **Douglas Hubbard** *is president and founder of Hubbard Decision Research, Chicago, Illinois, USA.* **Norman Fenton** *is professor of risk management at Queen Mary, University of London, UK, and chief executive of Agena, Cambridge, UK.* **Karl Claxton** *is professor in the Department of Economics and Related Studies, and senior research fellow in the Centre for Health Economics at the University of York, UK.* **Eike Luedeling** *is a senior decision analyst at the World Agroforestry Centre, Nairobi, Kenya.* **Jan de Leeuw** *is a drylands scientist at the World Agroforestry Centre, Nairobi, Kenya.*
e-mail: k.shepherd@cgiar.org

1. Sustainable Development Solutions Network. *Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for the SDGs* (SDSN, 2015).
2. Unterhalter, E. *J. Hum. Dev. Capab.* **15,** 176–187 (2014).
3. Seddon, J. *Systems Thinking in the Public Sector* (Triachy Press, 2008).
4. Sustainable Development Solutions Network. *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development* (SDSN, 2015).
5. Hubbard, D. W. *How to Measure Anything. Finding the Value of "Intangibles" in Business* (John Wiley, 2010).
6. Fenton, N. & Neil, M. *Risk Assessment and Decision Analysis with Bayesian Networks* (CRC Press, 2012).
7. Clapp, A., DauSchmidt, N., Millar, M., Hubbard, D. & Shepherd, K. *A Survey and Analysis of the Data Requirements for Stakeholders in African Agriculture* (World Agroforestry Centre, 2013).
8. Erisman, J. W. *et al. Nature* **519,** 151–153 (2015).
9. Howard, R. A. & Abbas, A. E. *Foundations of Decision Analysis* (Pearson, 2015).
10. Claxton K., Sculpher, M. & Briggs, A. *Decision Modelling for Health Economic Valuation* (Oxford Univ. Press, 2006).
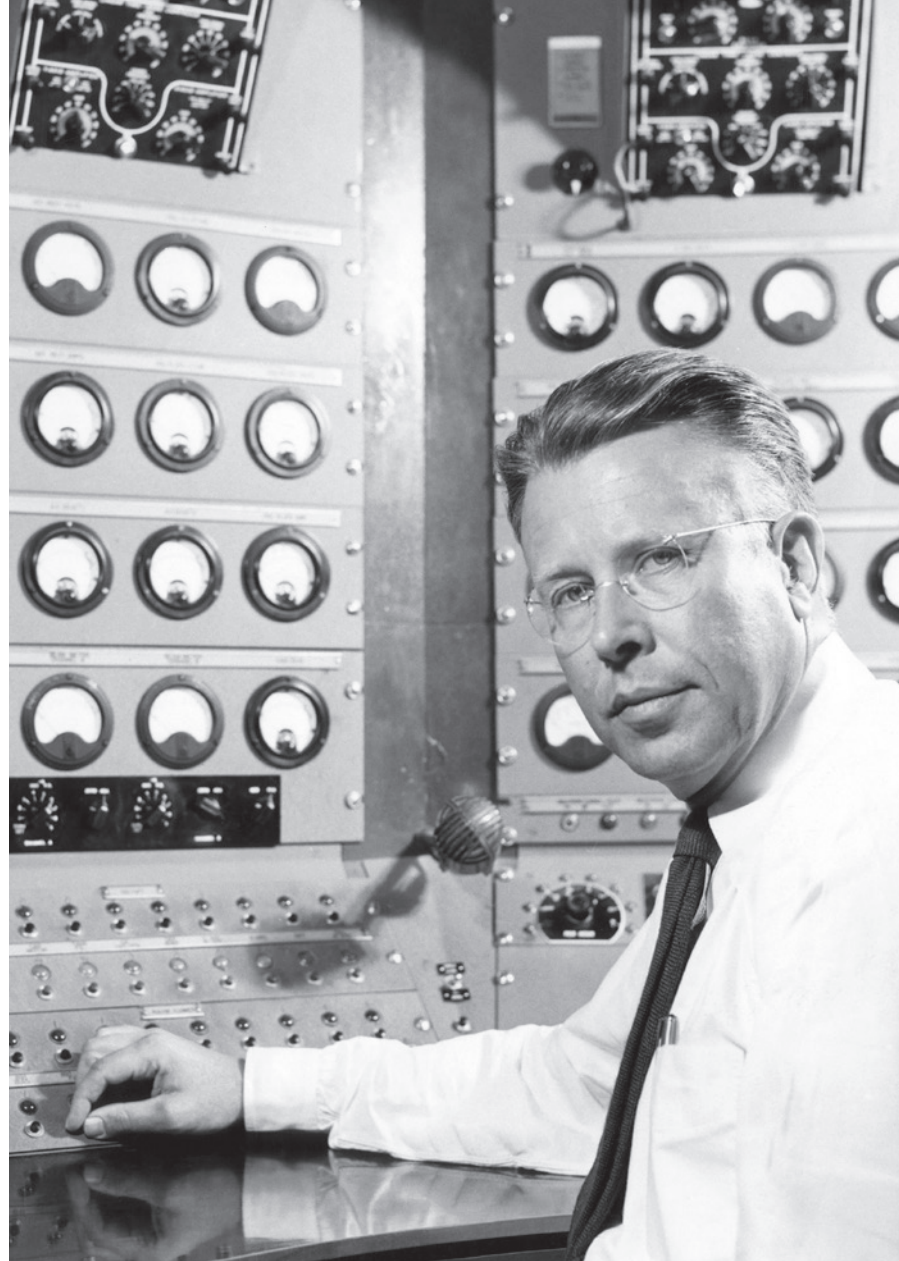
Ernest Lawrence co-invented the cyclotron particle accelerator.

Big Science: Ernest Lawrence and the Invention that Launched the Military-Industrial Complex
MICHAEL HILTZIK
Simon & Schuster: 2015.

PARTICLE PHYSICS

Inside the Rad Lab

Jon Butterworth relishes a tome on the research and the personalities that drove a century of smashing physics.

I approached Michael Hiltzik's *Big Science* with trepidation. I work on the biggest particle accelerator ever built — the Large Hadron Collider, which features in this tale — but the book looked heavy, at least in the gravitational sense, and I am not a fan of hagiographies. However, I was soon gripped. This is an astonishing story: US physicist Ernest Lawrence is at its core, but its scope is broad and full of context and characters.

Big Science spans the development of particle accelerators, the emergence of a team-driven approach to research, the

beginning of serious large-scale military, industrial and government sponsorship of science and the inception of fission and fusion weapons. Sometimes Lawrence is a cipher, cynically jumping from one funding source to another. At others he is a visionary leader of teams of genius, or an overstretched human being whose judgement and health eventually fail him.

The cyclotron, co-invented by Lawrence in 1932, takes advantage of the relationship between the magnetic force needed to bend the path of a particle in a circle, the radius of

the circle and the speed of the particle, to whizz particles around a loop and accelerate them to formerly unattainable velocities. Previously, physicists relied on natural sources to smash atoms. Ernest Rutherford's scattering experiment, which gave us the first look inside the atom, used α-particles emitted by radioactive radium. Other discoveries, such as that of the muon, were made by observing high-energy particles that bombarded Earth from space. The cyclotron offered beams much more intense than those from space, and of vastly higher energy than those from radioactive decay.

The ever larger and more efficient machines at Lawrence's Radiation Laboratory in Berkeley, California — the 'Rad Lab' — from 1931 onwards provided a bonanza of finds. Elements discovered there (such as lawrencium) carry the names of Rad Lab scientists. The group also supplied labs worldwide with isotopes for use in medicine.

The tensions between these sometimes conflicting priorities are convincingly described, as are Lawrence's methods of research management. When governments hymn impact and interdisciplinarity, they must surely hold the Rad Lab as a Platonic ideal. It is hard to beat the impact of developing the method of enriching uranium for the first atomic bomb. And Lawrence's team of obsessive scientists and engineers could be the definition of interdisciplinarity.

That team is the prototype for numerous cultural references. In Terry Pratchett's Discworld novel series, the High Energy Magic building at the wizards' Unseen University references the 1930s Rad Lab, with its camp beds, coffee and high-voltage atomsmashing. Lawrence's physician brother John bringing mice to be irradiated, and kicking off hadron cancer therapy; the violet deuteron beam used to impress visitors; the electromagnetic noise that meant that a light bulb pressed against any piece of copper piping would light up: all are examples of the potential for breakthrough and disaster. It is a world away from the Ivy League heights of US academia, or the "small science" citadels of Europe — Cambridge, Copenhagen, Göttingen and Manchester, which led physics into the quantum era but hungered for the technologies being born in Berkeley.

The growth of the Rad Lab's reputation ▶

and of US physics in general are well narrated. The newcomers made mistakes and missed opportunities, but European physicists — including such giants as Rutherford, James Chadwick and Pierre and Marie Curie — maintained a dialogue, and respected them. Lawrence and his team stayed engaged, increasingly willing to admit to errors as their confidence grew, and generous with their know-how in helping to start other accelerator programmes as the 'Cyclotron Republic' grew.

Compelling characters abound. There is the mysterious and influential Alfred Loomis, a patron of science who achieves the feat of "being a public figure without letting the public in on it". Later, there is Lewis Strauss (pronounced 'Straws'), Washington DC insider, chair of the Atomic Energy Commission and die-hard opponent of a nuclear-test-ban treaty. Lawrence seems to have easily formed bonds with exceptional people, but these sometimes shattered, as with Manhattan Project leader J. Robert Oppenheimer, causing damage and dismay.

Lawrence transformed strikingly from a man who insisted that politics had no place in the lab to one who played high-stakes political games around the credibility of scientific advice on nuclear-weapon development — and fired outstanding scientists because they refused to sign an oath of loyalty. The Rad Lab drew talent, but much of it leaked or was driven away as Berkeley became identified with the anti-communist McCarthyism — under which people were branded un-American and unemployable — that abounded in the military–industrial complex that it had helped to create.

The final chapter rushes through the formation of CERN in Geneva, Switzerland, and the failure of its US competitor, the Superconducting Super Collider, which was cancelled in the 1990s. It is a compliment to Hiltzik that, having initially worried about the book's size, I wanted more — in particular, on how CERN consciously distanced itself from the military aspect of the complex, and how the teamwork that Lawrence developed applies, or fails to, in collaborations of thousands rather than dozens. Lawrence had left the scene by then, but his influence still pervades academia, industry and politics. ∎

> **"The Rad Lab drew talent, but much of it leaked or was driven away as Berkeley became identified with McCarthyism."**

Jon Butterworth is professor of physics at University College London and writes for The Guardian at go.nature.com/qhea9i. He is the author of Smashing Physics. e-mail: j.butterworth@ucl.ac.uk

---

PHYSICS

The impulse of beauty

Joseph Silk revels in Frank Wilczek's treatise on how symmetry and harmony drive the progress of science.

Through the mythological figure of Urizen, William Blake probed the nature of reductionism.

Can beautiful ideas drive science? In *A Beautiful Question*, physicist and Nobel laureate Frank Wilczek makes a potent case that they can, hinging on qualities that have served as pathfinders to empirical truth in the physical world. The greatest scientists, from Galileo to Albert Einstein, saw in physics almost infinite beauty, including symmetry, harmony and truth. Today, we fervently hope for a genius with a beauty-inspired Theory of Everything — or at least for the Large Hadron Collider at CERN in Geneva, Switzerland, to discover truth in supersymmetry.

A Beautiful Question is both a brilliant exploration of largely uncharted territories and a refreshingly idiosyncratic guide to developments in particle physics. Vast and eclectic, it covers everything from atomism to the Higgs boson, musical harmony to anamorphic art, dark matter to the origins of the Universe. Wilczek lays out a vision of truth and beauty inspired by great modern physicists and classical philosophers such as Pythagoras and Plato. Lavish illustrations exemplifying beauty in art and science, from William Blake's *Ancient of Days* to fractal images, are interwoven with quotations from luminaries in the arts and sciences, from Molière to John Archibald Wheeler.

Wilczek begins with the beauty-inspired seeds sown by the ancient Greeks, including the fundamentals of geometry, music and chemistry. The music of the spheres, which Pythagoras described as the hum from celestial bodies whose periodicities echoed a harmony that he alone heard, inspired him and his followers to develop harmonies between beauty, music, mathematics and science. Numbers governed all, from octaves to right-angled triangles. Through perspective, geometry revolutionized classical, then Renaissance, art; through the curvature of space, it revolutionized understanding of gravity. And Wilczek argues that colour, the epicentre of beauty, unites art with biology, chemistry and physics.

The search for symmetry generated enormous rewards in science, a gift that has kept on giving. In the nineteenth century, Michael Faraday gave an elegant display of empirical physics by mapping out the patterns of magnetic lines of force. He went on to show that moving magnetic fields generate electric fields, motivating mathematical physicist James Clerk Maxwell to develop his equations for electromagnetism. These epitomized a fundamental symmetry, allowing a magnetic field in motion to generate an electric field, and vice versa. The fields propagate through space, producing waves of light in all colours of the rainbow. Maxwell's equations also predicted that electromagnetic waves would propagate at frequencies beyond perception by the human eye. Inspired, Heinrich Hertz discovered radio waves. Beauty had succeeded far beyond any intent of Faraday's.

Wielding the sword of beauty to refine scientific thought has a remarkable heritage. Einstein put beauty first in conceptualizing the general theory of relativity. In the dreary postwar climate of 1919, worldwide headlines greeted the successful verification of one of his key predictions — the bending of light by gravity. Another triumph is the standard model of particle physics, whose symmetries led to prediction of the Higgs boson.

Wilczek argues that the quantum core of modern physics, the zoo of elementary
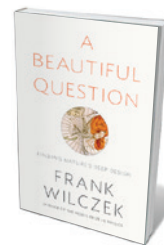
A Beautiful Question: Finding Nature's Deep Design
FRANK WILCZEK
Allen Lane: 2015.

particles, stems from beautiful thoughts framed by appeals to symmetry. The Eightfold Way, named by physicist Murray Gell-Mann after the Noble Eightfold Path of Buddhism, organizes elementary particles into octets; the Higgs, discovered in 2012, is the final missing link in the standard model.

Now the search is on for a unifying principle to take us back to simplicity. Supersymmetry, the most beautiful idea of all, unites two fundamental types of particles, fermions and bosons, distinguished by their spins. It postulates massive 'superpartners' for each particle, the lightest of which is a stable candidate for dark matter. Some see a lack of elegance in a theory that has some 120 adjustable degrees of freedom. The situation is, however, being redeemed in part through the enormous efforts of experimental particle physicists to measure many of these numbers. Only one real issue remains: at what energies must one smash particles together to seek supersymmetry's elusive signature? Wilczek optimistically predicts that we will discover this holy grail of physics in five years.

Occasionally the search for beauty has led us astray. Science was set back for centuries by the epicycles with which Greek astronomer Ptolemy described planetary motions. Modern data debunked Fred Hoyle's steady-state theory of the Universe. And even particle physics, with its grand hopes of unification, offers no insight into serious cosmological problems such as why dark matter is more than five times as abundant as ordinary matter. Most recently there has been string theory, the compellingly beautiful union of mathematical simplicity with quantum theory, particle physics and gravity. Its advocates have provoked a controversy: can a theory be so beautiful that we award it scientific accolades for its synthetic capacity without an empirical test, or must we dump it on the scrap heap of history for its lack of grounding truth?

Persistent voices insist that a theory of physics must lead to experimental verification. Wilczek is emphatic about this, as was Isaac Newton, who would like us to see empiricism as the search for truth. If truth and beauty are inseparable, that circle is closed. That is where supersymmetry will rise or fall. I hope for the latter, although I am reconciled to waiting for a new generation of unprecedentedly powerful particle colliders to reach the frontiers of our unifying theory. ∎

Joseph Silk *is at the Institut d'Astrophysique de Paris, the Beecroft Institute for Particle Astrophysics and Cosmology in Oxford, UK, and the Johns Hopkins University in Baltimore, Maryland.*
e-mail: joseph.silk@physics.ox.ac.uk

Books in brief
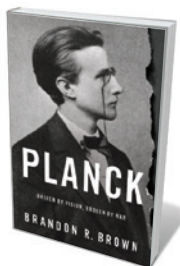
Planck: Driven by Vision, Broken by War
Brandon R. Brown OXFORD UNIVERSITY PRESS *(2015)*
The life of Max Planck, 'father of quantum theory', smacks of enigma: his personal papers were mostly destroyed in the Second World War. Physicist Brandon Brown has mined what survived for this illuminating biography. The main thread is the endgame of the Second World War, when the elderly Planck endured tribulations such as his son Erwin's trial and execution for treason against the Reich. Through this Brown interweaves a gripping backstory, ranging from Planck's landmark theoretical description of black-body radiation to his loyal advocacy for fellow physicist Lise Meitner.
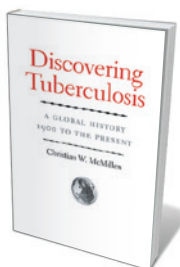
Discovering Tuberculosis: A Global History, 1900 to the Present
Christian W. McMillen YALE UNIVERSITY PRESS *(2015)*
Polio incidence is down by 99% since 1988, but tuberculosis (TB) remains a scourge; it kills 2 million people a year, most with HIV/AIDS. In his chronicle of TB's trajectory from the start of the twentieth century, historian Christian McMillen probes our failure to control this "resilient, powerful, protean bacterial infection" and its drug-resistant strains. Tracing the swathe TB has cut through Africa, India and Native American areas, McMillen identifies the catalogue of errors keeping it in circulation — such as the closure of the UK Medical Research Council's TB units in 1986, just as Africa's struggle with HIV began.

Secret Science: A Century of Poison Warfare and Human Experiments
Ulf Schmidt OXFORD UNIVERSITY PRESS *(2015)*
This monumental history of twentieth-century military medical ethics is a meticulous record of ambiguity. Historian Ulf Schmidt shows how Germany's use of chemical weapons such as mustard gas in the First World War spurred Britain, Canada and the United States to begin secret toxic-agent trials that purported, in some cases, to be benign medical testing. At the UK Porton Down research centre alone, Schmidt reveals, 21,752 soldiers took part in tests between 1939 and 1989 — an experience that was frequently unpleasant, occasionally harmful and in a few cases fatal.
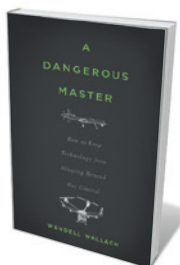
A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control
Wendell Wallach BASIC *(2015)*
Hordes of technologies emerge in lockstep with warnings of their risks. Ethicist Wendell Wallach sorts the hysteria from the hazards in this magisterial study. He looks in turn at disruption, complex systems, problematic trade-offs, the "transhumanism" movement — and new forms of governance to guide us through the innovatory onrush. It is conscious engagement, Wallach argues, that will allow us to resist the truly dangerous developments that threaten to "woo us to sleepwalk into the technological wonderland".
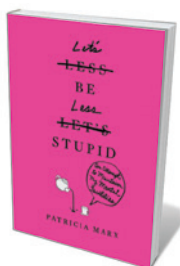
Let's Be Less Stupid: An Attempt to Maintain My Mental Faculties
Patricia Marx TWELVE *(2015)*
Struggling with brain fog? This "sub-primer" on the neuroscience of intelligence and memory by *New Yorker* staff writer and master humorist Patricia Marx delivers salutary cognitive jolts amid the general hilarity. Through a "higgledy-piggledy assortment of highfalutin science, lowfalutin sciences, tests" and more, Marx explores memory slippage, mindfulness, the Cherokee language and brain scans. If you regularly arrive in rooms with no memory of what you were looking for, this one is for you. **Barbara Kiser**

Correspondence

Standards needed for gene–editing errors

It is important to develop consensus guidelines for defining off-target mutations in DNA, which could occur as an unintended side-effect of genome editing (see *Nature* **522**, 20–24; 2015). I encourage the community to contribute to these discussions (see go.nature.com/zncbil).

Such uniform standards would help researchers, peer-reviewers, journal editors and regulators to best identify such mutations.

For therapeutic applications, unwanted mutations need to be defined by the most highly sensitive, unbiased genome-wide methods — given that even low-frequency events in large populations of cells could have clinical consequences. Such a comprehensive definition might not be necessary for research projects because appropriate control experiments would exclude the potentially confounding effects of off-target actions.

For now, direct comparison of state-of-the-art technologies can start to define best practices. Refinement will follow as detection and editing methodologies advance.
J. Keith Joung *Massachusetts General Hospital, Charlestown, Massachusetts, USA.*
jjoung@partners.org
Competing financial interests declared: see go.nature.com/dellef

Fieldwork grants would up diversity

You flag a social media debate on the practice of using volunteers for field research in biology (*Nature* **522**, 131; 2015). In our view, the solution is not to eliminate these positions, but to make them more worthwhile and accessible to students who need such experience — for example, to add weight to their graduate school applications or to test their commitment to field-research careers.

Unpaid internships are seen as elitist in that they can be taken up only by people who can afford to support themselves. However, banning volunteers would markedly reduce the availability of field positions.

A better strategy would be for funding agencies such as the US National Science Foundation to allocate research fellowships to trainees from under-represented socio-economic groups. This support would cover field expenses and provide a reasonable stipend. Such a system would also allow researchers to identify and recruit promising candidates and would facilitate valuable field experience for a more diverse set of applicants.
Joan B. Silk *Arizona State University, Tempe, Arizona, USA.*
joansilk@gmail.com
*On behalf of 4 correspondents (see go.nature.com/3e7y7v for full list).

Bolivia set to violate its protected areas

The Bolivian government has issued a decree allowing hydrocarbon exploration inside the country's protected areas. They have also given the green light for the construction of a controversial highway across the Isiboro Secure National Park and Indigenous Territory (TIPNIS). As scientists working in South American forests, we are concerned that these political developments override the country's international commitments and undermine the conservation of its unique biological and cultural diversity.

Several national and international groups, including activists and scientists, have voiced their opposition. The conflict has now reached a crucial stage, with President Evo Morales, once known as Bolivia's foremost defender of *Pachamama* ('Andean Earth Mother'), threatening to expel any non-governmental organization or foundation that attempts to obstruct the exploitation of the country's natural resources.

We call on the country's recently re-elected government to reconsider its environmental policies and to revisit its conservation pledges. We also urge the president to respect and support the legitimate and essential work of Bolivian civil organizations and their international partners in defending *Pachamama*.
Álvaro Fernández-Llamazares *Autonomous University of Barcelona, Spain; and University of Helsinki, Finland.*
Ricardo Rocha *University of Lisbon, Portugal; and University of Helsinki, Finland.*
Alvaro.FernandezLlamazares@uab.es

China should come clean on emissions

Uncertainties surrounding China's data on carbon emissions threaten to undermine its pledge for a 2030 emissions peak (see Z. Liu *et al. Nature* **522**, 279–281; 2015) and to confuse global strategies for preventing catastrophic climate change. The stakes are high: even small upward tweaks in China's coal consumption could generate more carbon dioxide than many countries emit in an entire year.

Emissions data are a sensitive issue in China, with official government statistical reports focusing more on energy production and consumption than on the country's binding carbon goals. China's carbon data are available for only 1994 and 2005 (through the UN Framework Convention on Climate Change) and are now outdated. Despite high-level policies in 2007 mandating a national greenhouse-gas statistical monitoring system, this has yet to materialize.

China says it reduced its CO_2 emissions per unit gross domestic product by 28.5% from 2005 to 2013, but our investigations suggest that more data are needed to confirm this. Clearer justification and methodological explanation is also needed for the frequent revisions of energy statistics by the National Bureau of Statistics of China in Beijing.

As Zhu Liu and colleagues point out, reliable monitoring systems and transparent reporting mechanisms are essential for China's internal emissions management.
Angel Hsu, Kaiyang Xu, Andrew Moffat *Yale School of Forestry and Environmental Studies, New Haven, Connecticut, USA.*
angel.hsu@yale.edu

TNF trailblazers five centuries apart

This year marks the 40th anniversary of a landmark paper describing the discovery of tumour necrosis factor (TNF), a pivotal cell-signalling protein in inflammatory disease known as a cytokine (E. A. Carswell *et al. Proc. Natl Acad. Sci. USA* **72**, 3666–3670; 1975). More than 122,000 publications on TNF followed — including reports that led to an important drug for treating arthritis, etanercept.

TNF's eponymous anti-cancer effects were unwittingly exploited by William Coley and colleagues as long ago as the end of the nineteenth century, after the likely induction of TNF by a mix of bacterial toxins (see B. Wiemann and C. O. Starnes *Pharmacol. Ther.* **64**, 529–564; 1994).

And many centuries earlier, in 1322, a Parisian midwife called Jacoba Felicie successfully burned the tissue around tumours to make them regress. We now know that burns cause inflammation, which activates TNF. As a woman, she was not permitted to qualify as a doctor, so she was put on trial for practising medicine and banished from Paris.
Claude Libert *Inflammation Research Center, VIB/University of Ghent, Ghent, Belgium.*
claude.libert@irc.vib-ugent.be

BIOGEOCHEMISTRY

Iron's voyage from the abyss

An iron–rich plume of water from a hydrothermal vent has been found to extend more than 4,000 kilometres through the ocean. The finding has implications for the productivity of marine algae, and therefore for climate. SEE LETTER P.200

KAZUHIRO MISUMI

In sunlit surface waters of the ocean, planktonic algae convert carbon dioxide from seawater into organic matter, which subsequently settles to the deep sea and sequesters the carbon from the atmosphere. In ocean regions that govern atmospheric CO_2 levels, the availability of iron — a trace nutrient — limits this primary production by algae[1]. Changes in iron availability have partly modulated climate variability during transitions from glacial to interglacial periods in the past[2], and are expected to affect future climate. On page 200, Resing et al.[3] report that a substantial amount of iron released from fissures in an abyssal mid-ocean ridge is transported thousands of kilometres by slow-moving deep-ocean currents. Using an ocean model, they show that iron from the global mid-ocean ridge is supplied to the euphotic layer (the sunlit surface region), and potentially contributes to algal growth.

In the late 1970s, scientists discovered hot, mineral-rich waters seeping from cracks in the sea floor called hydrothermal vents. Chemical analyses[4] of these waters showed them to be remarkably iron-rich compared with the surrounding ocean water. Hydrothermal iron was thought for decades to make only a minor contribution to the iron budget of the global ocean, because scientists assumed that it forms a solid precipitate near the vent sites as a result of its low solubility in seawater. Subsequent observations[5,6] showed, however, that some of the iron released from hydrothermal vents may be transported away from vent sites. This possibility is called the leaky vent hypothesis[7].

Scientists have detected unusually high concentrations of dissolved iron distributed over horizontal distances of hundreds to thousands of kilometres in various deep ocean basins[8–11]. The hydrothermal origin of these iron-rich anomalies has been inferred by examining the isotopic signatures of helium within them. Helium has two stable isotopes, ^{3}He and ^{4}He, and hydrothermal water is enriched in ^{3}He relative to the proportion found in atmospheric helium — there is said to be excess ^{3}He. The presence of excess ^{3}He therefore indicates water of hydrothermal origin[10]. Correlations between the anomalous dissolved
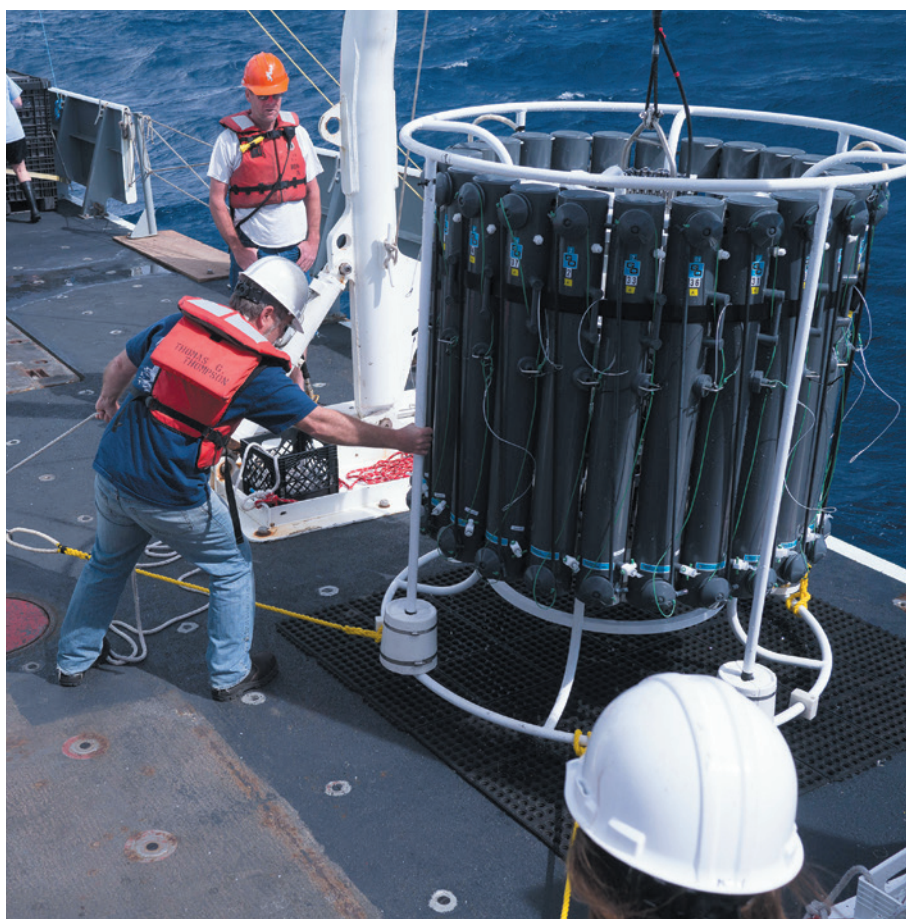
Figure 1 | Sampling seawater. By collecting hundreds of seawater samples around the southern part of a mid-ocean ridge called the East Pacific Rise, Resing et al.[3] discovered an iron-rich plume of water that extends for more than 4,000 kilometres from a hydrothermal vent. Here, a sampler is recovered after collecting water from the deep ocean.

iron concentrations and excess ^{3}He have been reported[10,12], but the data for helium and for dissolved iron were collected at different times, and so the hydrothermal origin of the dissolved-iron anomalies could not be confirmed.

Resing and colleagues collected hundreds of seawater samples (Fig. 1) around the southern part of a mid-ocean ridge called the East Pacific Rise, which is particularly active volcanically. They identified a remarkable plume of water containing high concentrations of dissolved iron extending more than 4,000 km downstream. The dissolved-iron concentration correlated linearly with the concentration of excess ^{3}He, which was measured using simultaneously collected helium data. This unambiguously proves that the anomalous dissolved iron is derived from hydrothermal vents in the East Pacific Rise.

Furthermore, the authors found that the linear relationship between dissolved iron and excess ^{3}He was maintained throughout the large plume, which indicates that the same process — inferred to be dilution with the surrounding seawater — controls the concentration of both helium and dissolved iron. Such conservative behaviour is inconsistent with the expected chemical behaviour of inorganic

iron, and supports an aspect of the leaky vent hypothesis: the idea that physico-chemical stabilization enables iron to be transported long distances from hydrothermal vent sites. Moreover, from the linear relationship, the authors estimated that the amount of iron transported globally from hydrothermal vents is 3–4 gigamoles per year, which is more than 4 times higher than previous estimates.

Resing and co-workers went on to use a cutting-edge global ocean model[13] to estimate the contribution of hydrothermal iron to the export of organic carbon from the euphotic layer, and found that the contribution was substantial, especially in the Southern Ocean. This has implications for the role of hydrothermal iron in past, present and future climates. For example, during glacial periods, increased deposition of iron-bearing dust onto the ocean surface is thought to have contributed to the lowering of atmospheric CO_2 concentrations[2]. But a stable supply of hydrothermal iron over millennial timescales would have buffered short-term variations of the iron supply, and thus also of oceanic CO_2 uptake, casting this theory in doubt.

Many questions need to be answered before the role of hydrothermal iron in marine biogeochemical cycles can be fully understood. One issue is that the size of global hydrothermal iron flux is highly uncertain. Resing *et al.* estimated the global flux using data from a single hydrothermal system, but the relationship between levels of iron and ^{3}He is likely to vary for different hydrothermal sites, because the tectonic history and the chemical compositions of the surrounding rocks will differ. For example, the ratio of the concentration of dissolved iron to that of ^{3}He is 80-fold higher in the southern Atlantic Ocean than in the southern Pacific Ocean[10]. More data are therefore needed from different sites to constrain estimates of the global hydrothermal iron flux, and the mechanisms causing variability among sites must be better understood.

Another issue concerns the mechanism by which iron is stabilized around hydrothermal vents. One possible mechanism is the formation of complexes between iron ions and organic ligand molecules[5]. Organic ligands are thought to be ubiquitous in seawater and to control dissolved iron concentrations by increasing iron solubility — more than 99% of dissolved iron in seawater is organically complexed[14]. But our knowledge of the sources and sinks of organic ligands in the ocean is still limited, and most global ocean models assume a fixed ligand concentration. The global ocean model[13] used by Resing and colleagues mechanistically represents the dynamics of organic ligands in the ocean. The team could thus simulate the transport of organically complexed iron away from the hydrothermal vents. Although the model makes several assumptions, the authors' results highlight the value of mechanistic representations of ligand dynamics in such models.

Will hydrothermal iron continue to be a nutrient source for surface algae? Rapid environmental changes have been occurring in the Southern Ocean over the past few decades: increases in the levels of greenhouse gases and the depletion of stratospheric ozone have led to intensified upwelling of deep water[15], whereas recent amplification of the water cycle and the melting of Antarctic glaciers has strengthened surface-water stratification[16]. These changes have the potential to alter exchanges between surface and deep waters, and thus the contribution of hydrothermal iron to surface biological productivity. ■

Kazuhiro Misumi *is at the Environmental Science Research Laboratory, Central Research Institute of Electric Power Industry, Abiko, Chiba 270-1194, Japan.*
e-mail: misumi@criepi.denken.or.jp

1. Moore, C. M. *et al. Nature Geosci.* **6,** 701–710 (2013).
2. Kohfeld, K. & Ridgwell, A. in *Surface Ocean: Lower Atmosphere Processes* (eds Le Quéré, C. & Saltzman, E. S.) 251–286 (Am. Geophys. Un., 2009).
3. Resing, J. A. *et al. Nature* **523,** 200–203 (2015).
4. Von Damm, K. L. *Annu. Rev. Earth Planet. Sci.* **18,** 173–204 (1990).
5. Bennett, S. A. *et al. Earth Planet. Sci. Lett.* **270,** 157–167 (2008).
6. Yücel, M., Gartman, A., Chan, C. S. & Luther, G. W. III *Nature Geosci.* **4,** 367–371 (2011).
7. Toner, B. M., Marcus, M. A., Edwards, K. J., Rouxel, O. & German, C. R. *Oceanography* **25,** 209–212 (2012).
8. Wu, J., Wells, M. L. & Rember, R. *Geochim. Cosmochim. Acta* **75,** 460–468 (2011).
9. Nishioka, J., Obata, H. & Tsumune, D. *Earth Planet. Sci. Lett.* **361,** 26–33 (2013).
10. Saito, M. A. *et al. Nature Geosci.* **6,** 775–779 (2013).
11. Conway, T. M. & John, S. G. *Nature* **511,** 212–215 (2014).
12. Fitzsimmons, J. N., Boyle, E. A. & Jenkins, W. J. *Proc. Natl Acad. Sci. USA* **111,** 16654–16661 (2014).
13. Völker, C. & Tagliabue, A. *Mar. Chem.* **173,** 67–77 (2015).
14. Rue, E. L. & Bruland, K. W. *Mar. Chem.* **50,** 117–138 (1995).
15. Lenton, A. *et al. Geophys. Res. Lett.* **36,** L12606 (2009).
16. de Lavergne, C., Palter, J. B., Galbraith, E. D., Bernardello, R. & Marinov, I. *Nature Clim. Change* **4,** 278–282 (2014).

CANCER

Diagnosis by extracellular vesicles

The detection of a single molecule anchored to circulating extracellular vesicles allows late–stage pancreatic cancer to be identified from just one drop of a patient's blood. SEE ARTICLE P.177

CLOTILDE THÉRY

On page 177 of this issue, Melo *et al.*[1] describe a non-invasive test that identifies patients with late-stage pancreatic cancer with 100% certainty, and that can distinguish patients with precancerous pancreatic lesions from those with benign pancreatic diseases. Although the number of patients in the precancerous-lesion group was low, and the findings require further validation in a larger cohort, the potential implications of such a test are huge. It would allow clinicians to decide whether or not to perform potentially debilitating surgery.

The test involves detecting a membrane-anchored proteoglycan molecule, glypican-1 (GPC1), in vesicles that circulate in the bloodstream. The authors found this proteoglycan in membranous material isolated from a small amount of frozen serum taken from all tested patients who had pancreatic cancer. By contrast, the sera of patients with other pancreatic diseases did not contain higher levels of GPC1-containing (GPC1$^+$) vesicles than those of healthy donors (Fig. 1). The test was more reliable than a commonly used assay (which involves the ELISA method) to detect the presence in whole blood of a pancreatic-tumour biomarker called carbohydrate antigen 19-9 (CA 19-9). About half of the patients without cancer had elevated CA 19-9 levels, whereas none had elevated GPC1$^+$ vesicles, and CA 19-9 was not elevated above control levels in many patients with cancer. Furthermore, in a mouse model of genetically induced pancreatic cancer, Melo and colleagues' test gave positive results before a detectable tumour was present.

Overexpression of GPC1 in pancreatic carcinoma and a positive role for this overexpression in tumour proliferation and metastasis have been reported previously using tumour cell lines and mouse models[2–4]. The novelty of Melo and colleagues' report resides in the presence of GPC1 in circulating vesicles in serum, and in the striking value of this molecule as a biomarker. Note that simple detection of serum GPC1 by ELISA, without concentrating the vesicles, does not provide a more reliable diagnostic test than CA 19-9 detection. Thus, this work demonstrates
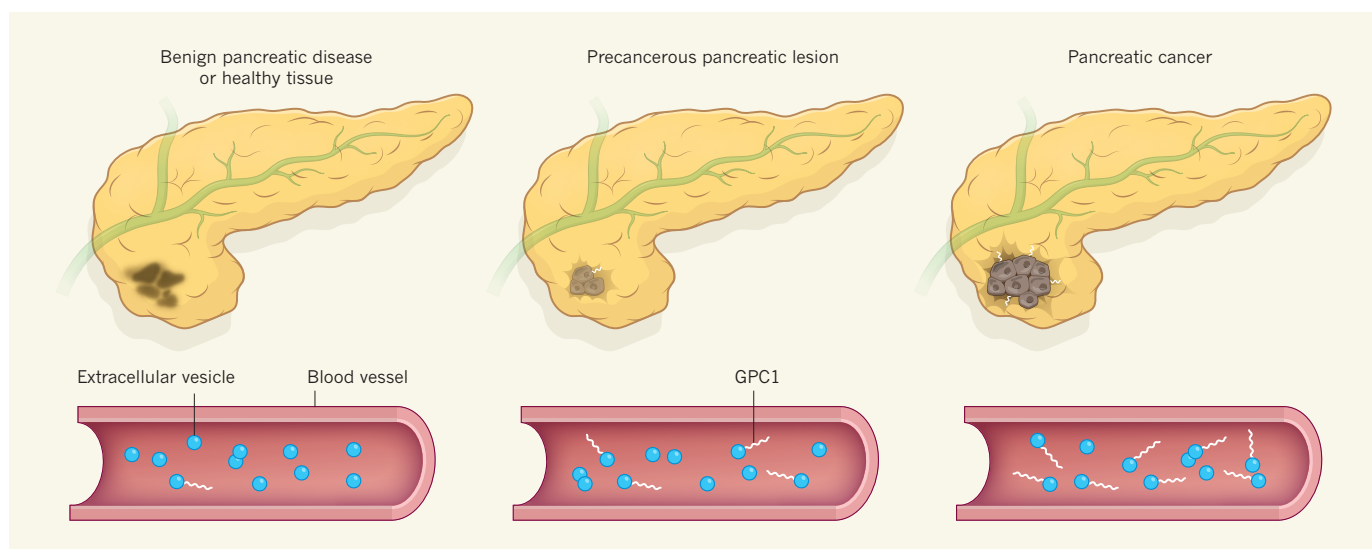
Figure 1 | GPC1 distinguishes pancreatic cancer from benign disease. Melo *et al.*[1] show that extracellular vesicles isolated from the bloodstream of patients with precancerous pancreatic lesions or pancreatic cancer contain the membrane-anchored proteoglycan GPC1, whereas this molecule is not found at higher-than-normal levels in vesicles in the blood of patients with non-cancerous pancreatic disease or of healthy donors.

for the first time that circulating vesicles in blood can be a source of specific and reliable diagnostic biomarkers for cancer.

Vesicles present in bodily fluids, which are collectively known as extracellular vesicles (EVs) or exosomes (as Melo and colleagues call them), have been investigated as potential biomarkers for various diseases for a decade[5–7]. But until now, overexpression of exosomes or exosomal markers was observed either only in advanced disease or after (rather than before) detectable tumour progression[8], or without statistical significance[9]. Recently published exosome analyses of the blood of patients with lung[10] or pancreatic cancer[11] have reported cancer detection with 75% and 93% specificity, respectively. However, the studies respectively measured the expression of a combination of 30 proteins on a microarray chip or a combination of 5 proteins and 4 microRNA molecules. These tests are less reliable and more complex than Melo and colleagues' test, which involves the detection of a single molecule and more-conventional techniques.

The authors' protocol uses long ultracentrifugation of small volumes of serum, coating of beads with the resulting pellet, and staining of the beads with a GPC1-specific antibody before analysing them by flow cytometry. Ultracentrifuges and flow cytometers are widespread and straightforward to use, suggesting that this protocol could be implemented in clinical laboratories as a routine procedure for evaluating patients who present with symptoms of pancreatic disease. The authors also show that the beads that have captured GPC1+ vesicles contain a mutant messenger RNA expressed by the tumour, which could allow further exploration of tumour properties.

On a slightly disappointing note, it seems that this test might not be useful for cancers other than pancreatic cancer. Although the authors' identification of GPC1 as a cancer-specific protein secreted in exosomes involved comparing cancerous and non-cancerous cell lines of breast origin, expression of GPC1 in circulating vesicles did not reliably identify patients with breast cancer, nor allow patients to be assigned to a specific breast-cancer subtype. However, GPC1 expression in breast cancer may still warrant further exploration. I noticed that among those patients whose blood EVs showed GPC1 expression, there were two distinct populations, with either a high or an intermediate number of GPC1+ EVs. The authors did not discuss this observation, but I wonder if the amount of GPC1+ circulating EVs could provide additional diagnostic or prognostic information.

Finally, I would like to devote a few words to the term exosome. It was first used in the context of vesicles in 1981 to describe membrane-enclosed structures of variable size (either 40 nanometres or 500–1,000 nm in diameter) that had been 'exfoliated' from the surface of cultured cells[12]. The same term was then proposed in 1987 for small (50–100 nm) vesicles that form inside cellular compartments called endosomes, and that are released extracellularly when these compartments fuse with the cell membrane[13]. Several research groups, including mine, have defended this latter use, but as EVs have become the focus of increasing interest, the word exosome has started to be used for small EVs without showing that they arose from endosomes rather than from the cell membrane[14].

The EVs used by Melo *et al.* for GPC1-based diagnostics are recovered from an untracentrifugation pellet that will contain exosomes as well as other types of small EVs, lipoproteins and even complexes of proteins and nucleic acids. Because GPC1 is a membrane-anchored protein, it is probably recovered in EVs, but the authors do not show the origin of the EVs that have diagnostic value. Melo and colleagues' paper will probably contribute to the increasing popularity of the term exosome, and possibly a generalization of its use to any type of small EV, about which purists such as myself can probably do little. That said, the intracellular origin of circulating GPC1 is irrelevant to its use in a diagnostic test, and a semantic issue should not interfere with the diffusion of such clinically important findings. ∎

Clotilde Théry *is in the Exosomes and Tumor Growth team, Immunity and Cancer Unit, Institut Curie, 75249 Paris Cedex O5, France. e-mail: clotilde.thery@curie.fr*

1. Melo, S. A. *et al. Nature* **523,** 177–182 (2015).
2. Kleeff, J. *et al. J. Clin. Invest.* **102,** 1662–1673 (1998).
3. Aikawa, T. *et al. J. Clin. Invest.* **118,** 89–99 (2008).
4. Whipple, C. A., Young, A. L. & Korc, M. *Oncogene* **31,** 2535–2544 (2012).
5. Kim, J. W. *et al. Clin. Cancer Res.* **11,** 1010–1020 (2005).
6. Zhou, H. *et al. Kidney Int.* **70,** 1847–1857 (2006).
7. Skog, J. *et al. Nature Cell Biol.* **10,** 1470–1476 (2008).
8. Peinado, H. *et al. Nature Med.* **18,** 883–891 (2012).
9. Logozzi, M. *et al. PLoS ONE* **4,** e5219 (2009).
10. Jakobsen, K. R. *et al. J. Extracell. Vesicles* **4,** 26659 (2015).
11. Madhavan, B. *et al. Int. J. Cancer* **136,** 2616–2627 (2015).
12. Trams, E. G., Lauter, C. J., Salem, N. Jr & Heine, U. *Biochim. Biophys. Acta* **645,** 63–70 (1981).
13. Johnstone, R. M., Adam, M., Hammond, J. R., Orr, L. & Turbide, C. *J. Biol. Chem.* **262,** 9412–9420 (1987).
14. Gould, S. J. & Raposo, G. *J. Extracell. Vesicles* **2,** 20389 (2013).

This article was published online on 24 June 2015.

BIODIVERSITY

Hidden impacts of logging

A meta-analysis of changes in the abundance of tropical-forest birds reveals that the effect of selective timber harvesting varies with logging practices and species traits. The results offer a framework for managing impacts on biodiversity.

JOSEPH A. TOBIAS

Tropical forests are famed for their exceptional biological richness, but the future of this biodiversity is increasingly threatened by land-use change. Selective logging — the commercial extraction of valuable timber species — is perhaps the most widespread and rapid form of change, currently affecting at least one-fifth of the remaining tropical forests and proceeding at 20 times the rate of clear-felling (full deforestation)[1]. The effect of selective logging on biodiversity has been the focus of intensive debate, fuelled by decades of local-scale studies that have generated contradictory results. Writing in *Proceedings of the Royal Society B*, Burivalova *et al.*[2] describe an ambitious attempt to resolve this issue by combining local-scale results into a pan-tropical meta-analysis of bird species responses to selective logging.

The authors quantified species responses as the difference in abundance or population density between a logged site and a nearby control site surveyed with the same methods. This focus on abundance is a forward step because broad-scale analyses are often forced to rely on 'presence or absence' data, thus concealing many of the effects of land-use change on biodiversity[3]. With more than 4,000 matched observations for a total of nearly 1,000 species, Burivalova and colleagues' data set has the statistical power and flexibility needed to unravel the complex effects of selective logging on bird populations. The authors capitalize on this potential by building models that simultaneously consider variations in both logging practices and species traits.

The main environmental variables considered were logging intensity, time since logging, the number of logging cycles and the type of logging undertaken. Type of logging ranged from sustainable forestry practices, such as reduced-impact logging, to more-damaging

conventional logging. Introducing this level of detail is valuable because variation in the history and type of logging can have markedly different effects on biodiversity[4,5]. Moreover, it is not enough to focus exclusively on the environment because the intrinsic ecological and life-history traits of different species also influence their responses to land-use change[6–8]. The authors therefore incorporated several such traits in their analyses, including diet and body mass, as well as differences in sensitivity to human pressures, such as hunting.

This analytical approach revealed that responses to logging vary according to species traits. For example, the feeding groups most adversely affected by logging were fruit-eating and insect-eating bird species, both of which declined in abundance at high logging intensity. By contrast, populations of nectar-eaters and seed-eaters increased in response to selective logging, at least until forest regeneration closed the tree canopy. This is not particularly surprising: many frugivorous and insectivorous species are forest specialists, whereas nectarivores and granivores are typically associated with non-forested or lightly forested habitats where flowering and seed-bearing plants are more abundant. Similarly, the authors' pan-tropical models revealed that different forms of logging practice had varying, but intuitive, outcomes — higher logging intensity caused the most lasting changes to bird populations (Fig. 1).

Earlier studies have reported similar patterns, both in the effects of logging practices[4,5] and in species traits[6–8]. However, Burivalova *et al.* were able to show that the best-fitting model of avian responses to logging does not pinpoint the dominant predictor to be either logging practices or species traits, but instead a combination of both. Their findings show, for example, that the most important predictors of shifts in abundance after logging were the time elapsed since the most recent logging event and the feeding group of the species involved, along with the interaction between these factors.

A key implication of these findings is that models incorporating information about planned logging practices and species traits could be used to predict the response of individual species, or communities of species, to future logging events. Similar models could even be applied retrospectively to evaluate the impact (and appropriate mitigation) of completed operations in areas where biodiversity was not monitored. Whether these applications are viable remains to be seen, because the models in their current form are relatively crude and have limited predictive power.

This is partly because the meta-analysis was based on a restricted sample of 26 studies, spanning a range of local contexts. Although Burivalova *et al.* made every effort to ensure comparability between these studies, many unavoidable inconsistencies remain — in

Figure 1 | Logged and loaded. Burivalova *et al.*[2] show that selective logging has long-lasting implications, with populations of some bird species showing little sign of recovery even 40 years after timber extraction.

50 Years Ago

In a written answer in the House of Commons on June 24, the Minister of Technology, Mr. F. Cousins, gave the names of 17 research associations which actively encouraged the use of computers in their respective industries; of 18 research associations which had access to computers on their premises, at universities or at member firms … In another written answer on June 24, Mr. Cousins stated that of 4,064 non-industrial Civil Servants employed by his Department … 1,400 had university degrees or equivalent qualifications in scientific or technological subjects, and about another 1,400 had other scientific or technological qualifications. In a third written answer, Mr Cousins stated … action was in hand … to promote the greater use of technological subjects in television and radio programmes, and to produce special booklets and films for wide distribution among young people.
From *Nature* **10 July 1965**

100 Years Ago

Among the recent additions to the zoological department at South Kensington are some specimens which are surely destined to possess historical interest for posterity. They consist only of two or three examples of harvest-mice and one house-mouse, but they were caught in the trenches in northern France, in that part of the trenches, in fact, occupied by some of our Indian troops. These specimens were collected and presented to the museum by one of the officers of an Indian regiment, whose keenness for his favourite pursuit of natural history allowed him in the intervals of being heavily shelled by the enemy a little relaxation in the way of trapping and skinning any animals for the national museum in London.
From *Nature* **8 July 1915**

factors such as spatial scale, logging-practice terminology, disturbance history, hunting pressure, road-building activity, survey intensity and observer experience. Moreover, although the total species list seems extensive, it contains numerous open-country or garden birds (such as the common bulbul *Pycnonotus barbatus* and the house wren *Troglodytes aedon*), along with highly conspicuous dispersive taxa (such as parrots and raptors) that may have been observed flying between primary forest patches rather than using logged forests. Inclusion of these categories may obscure the key impacts of logging on populations of forest-dependent species. Similar issues arise with species traits, which Burivalova *et al.* treat in a simplified form. For example, the authors assigned bird species to one of seven feeding groups (carnivores, insectivores, granivores, nectarivores, frugivores, omnivores or herbivores), but many species belong in multiple categories, and shift between categories over space and time[9].

Many of these issues can be addressed by expanding or refining the underlying environmental and biological data. Attempts should be made to coordinate and standardize methods across the current spate of long-term initiatives that monitor the effects of selective logging at the local and landscape scale in tropical and temperate forests. In addition, the immediate prospects for improving information on species traits are good, particularly for birds. For instance, comprehensive data sets that describe the diet, habitat use and biometrics of birds are available (see ref. 9, for example). These offer a more nuanced assessment of key attributes such as dietary niche and dispersal ability, which are relevant to ecosystem processes such as seed dispersal.

Incorporating these advances into global models will shed further light on the role of species traits in predicting responses to land-use change, as well as the broader implications for ecosystem function and services[10,11]. Thus, although Burivalova and colleagues' efforts may fall short of providing a workable model for sustainable forestry, they point the way to more-sophisticated approaches that can help us to understand the impacts of selective logging on biodiversity, and to develop guidelines for logging practices that balance the needs of people with biodiversity across the tropics and beyond. ∎

Joseph A. Tobias *is in the Department of Life Sciences, Imperial College London, Silwood Park, Ascot SL5 7PY, UK.*
e-mail: j.tobias@imperial.ac.uk

1. Asner, G. P., Rudel, T. K., Aide, T. M., Defries, R. & Emerson, R. *Conserv. Biol.* **23,** 1386–1395 (2009).
2. Burivalova, Z. *et al. Proc. R. Soc. B* **282,** 20150164 (2015).
3. Bregman, T. P., Şekercioğlu, C. H. & Tobias, J. A. *Biol. Conserv.* **169,** 372–383 (2014).
4. Burivalova, Z., Şekercioğlu, C. H. & Koh, L. P. *Curr. Biol.* **24,** 1–6 (2014).
5. Bicknell, J. E., Struebig, M. J., Edwards, D. P. & Davies, Z. G. *Curr. Biol.* **24,** 1119–1120 (2014).
6. Newbold, T. *et al. Proc. R. Soc. B* **280,** 20122131 (2013).
7. Cleary, D. F. R. *et al. Ecol. Appl.* **17,** 1184–1197 (2007).
8. Hamer, K. C. *et al. Biol. Conserv.* **188,** 82–88 (2015).
9. Wilman, H. *et al. Ecology* **95,** 2027 (2014).
10. Edwards, D. P., Tobias, J. A., Sheil, D., Meijaard, E. & Laurance, W. F. *Trends Ecol. Evol.* **29,** 511–520 (2014).
11. Ewers, R. M. *et al. Nature Commun.* **6,** 6836 (2015).

A twist in the tale of γ-ray bursts

An unusually long burst of γ-rays zapped Earth in December 2011, lasting 4 hours. The cause of this burst is now proposed to be a peculiar supernova produced by a spinning magnetic neutron star. SEE LETTER P.189

STEPHEN J. SMARTT

The story of γ-ray bursts (GRBs) originates in nuclear-weapons monitoring during the cold war, and has been elaborated by subsequent technological developments and scientific detective work. GRBs were discovered by the Vela satellites launched in the late 1960s by the US Air Force. The spacecraft carried sensitive γ-ray detectors to monitor the Soviet Union's compliance with the Nuclear Test Ban Treaty. No nuclear explosions on Earth were seen. Instead, mysterious γ-ray flashes were detected, randomly distributed on the sky[1]. On page 189 of this issue, Greiner *et al.*[2] present data for a γ-ray flash that suggest an association with a rare type of supernova, similar to an unusual type of stellar explosion that has been recognized only in the past few years[3].

Nearly 50 years after the end of the cold war, following several space missions dedicated to high-energy astronomy and the harnessing of the most powerful ground-based telescopes,

we have a clearer picture of the cause of GRBs. They fall into two main types, defined simply by the duration of the γ-ray emission. Short GRBs last between about 0.1 and 1 second, whereas long GRBs last from about 2 seconds to several minutes. The leading model for the production of short GRBs is the merger of a black hole and a neutron star (the dense nucleus of a dead massive star), or of a pair of neutron stars. It is thought that these mergers might also produce bursts of gravitational waves. If these can be recorded by future detectors[4], it would directly validate one of general relativity's tenets and revolutionize physics.

Long GRBs are the more commonly detected, and make up about 70% of all the events detected by the Swift satellite, which is dedicated to the discovery of GRBs. The nearest known long GRB was located 40 million parsecs away, far beyond the local group of galaxies that contains the Milky Way. Relatively close events such as this tend to be accompanied by a particular type of supernova explosion. The leading hypothesis for their origin involves the collapse of a massive star followed by the formation of a black hole, which is surrounded by a spinning disk of gas assembled from the star's remains. As material from the disk falls onto the black hole, a jet is launched that accelerates particles close to the speed of light. The relativistic jet is focused in a beam and creates γ-ray emission, followed by X-ray and optical radiation. The signature of the supernova emerges days later, when the emission from the beam, known as the afterglow, fades rapidly.

In the past few years, a rare type of GRB has been discovered[5], known as an ultra-long GRB. As the name suggests, the duration of the γ-ray emission from these can be several thousand seconds, but the events seem to be of similar power (energy emitted per second) to the bulk of the known long and short GRBs (Fig. 1). The most likely cause for one of the detected ultra-long GRBs — which lasted several days — was the tidal (gravitational) disruption of a star as it was being consumed by a supermassive black hole at the centre of a galaxy[6]. However, there are three recorded GRBs for which neither the supernova link nor the cause of the tidal disruption could be established[5].

Greiner and colleagues studied the nearest of these, known as GRB 111209A. Its γ-ray emission lasted about 4 hours, and an afterglow detected at optical wavelengths allowed an accurate determination of its redshift (0.677). This tells us that the Universe was about half its current age when the light from
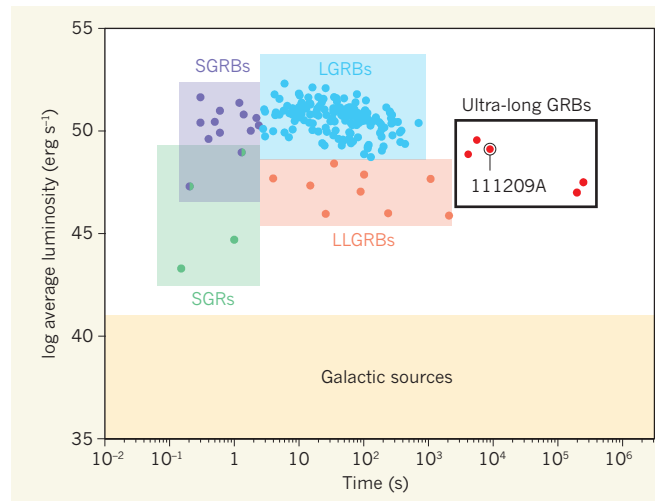
Figure 1 | Energetics and duration of γ-ray bursts. The astrophysical phenomena known as γ-ray bursts (GRBs) can be grouped according to their luminosity (energy emitted per second) and duration. The lowest-energy phenomena are located in our Galaxy and fall in the shaded region at the bottom of this diagram. Soft γ-ray repeaters (SGRs) are magnetic neutron stars also found in the Milky Way. Low-luminosity GRBs (LLGRBs) are probably a distinct class of extragalactic source. Most of the transient γ-ray flashes observed in the sky are the massively energetic short or long GRBs (SGRBs and LGRBs, respectively) located at huge cosmological distances. Their duration ranges from about a fraction of a second to many minutes, respectively. The ultra-long GRBs, which include the 4-hour-long GRB 111209A discussed by Greiner et al.[2], are distinctly different from the bulk of the population. (Adapted from ref. 5.)

the supernova was emitted. The unusual properties of 111209A have been studied by several groups[5,7,8], but Greiner and colleagues' deep optical and near-infrared imaging now provide a superior data set suitable for a study of the decaying emission from the explosion.

After forensically detailed examination of the GRB's light curve — which depicts the evolution of its luminosity over time — the authors observed a distinct bump. This bump is the signature of a luminous supernova, corresponding to an explosion that is several times brighter than those typically associated with long GRBs. The spectrum of this supernova is also atypical: it does not show the strong absorption features due to iron that are prominent in the spectra of events associated with long GRBs. The authors suggest that this spectrum is similar to those of a new class of super-luminous supernova discovered in 2011 (ref. 3). Never before has the super-luminous class been directly associated with GRBs.

Super-luminous supernovae are some 10–100 times brighter than all other types, and evolve slowly. Most of them cannot be powered by the radioactive decay of the nickel isotope ^{56}Ni, which is the standard physical model invoked to explain the light emitted from supernovae previously associated with GRBs. However, the light curves of super-luminous supernovae can be quantitatively modelled if extra energy is injected into the explosion

from a rapidly spinning neutron star[9]. Neutron stars are formed when massive stars collapse, and the release of gravitational potential energy is what causes a normal supernova. Neutron stars that are born with spin periods of less than about 10–20 milliseconds have enough energy to power the emission observed by the authors.

The physical mechanism through which rotational energy can be extracted from a rapidly spinning neutron star is the emission of radiation from the object's magnetic-field poles (magnetic dipole radiation); neutron stars known as magnetars, which have magnetic fields of around 10^{14} gauss, can lose[10,11] rotational energy through this mechanism on the timescales required to explain the authors' observations. These are extreme, but physically plausible, field strengths, and it has been shown that simple models of magnetic dipole radiation from magnetars do match the data from super-luminous supernovae quite well[12].

Greiner et al. have reached the secure conclusion that an unusual and luminous supernova accompanied the ultra-long GRB 111209A. Their deduction stretches the standard physical model invoked to explain supernova luminosities to breaking point. Although, in this case too, the simple magnetar models fit the light curve quite well, they contain several free parameters that are unconstrained. The magnetic-field strength, spin period and ejecta mass can be chosen to fit many shapes of supernova light curves, not just this one. The models have therefore been criticized for being too flexible. The current analysis cannot confirm that a magnetar is indeed the powering mechanism of the supernova. Also, the low signal-to-noise ratio of the spectrum obtained does not lend itself to an unambiguous conclusion.

This possible link between super-luminous supernovae and GRBs requires further investigation, but the quest is hampered by the rarity of both phenomena. Greiner and colleagues' supernova is only one example, and although its light curve and spectrum bear some resemblance to those of super-luminous supernovae, they are certainly not a perfect match. It is also intriguing that most GRBs and super-luminous supernovae have been found in low-mass dwarf galaxies[13]. Dwarf galaxies are likely to have lower abundances of elements heavier than helium than the galaxies in which the bulk of star formation occurs in the Universe[14]. The similarity of the birthplaces of GRBs and super-luminous supernovae has

also suggested[13] a link between them.

Fifty years after a military mission unexpectedly made one of the most remarkable discoveries in high-energy astronomy, we are still struggling to unify the physical models of GRBs. Greiner and co-workers' findings add another twist to the tale of γ-ray astronomy, which will undoubtedly be followed by others in the next few years, when gravitational-wave detectors start surveying high-energy phenomena in the sky. ■

Stephen J. Smartt *is at the Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK.*
e-mail: s.smartt@qub.ac.uk

1. Klebesadel, R. W., Strong, I. B. & Olson R. A. *Astrophys. J.* **182,** L85 (1973).
2. Greiner, J. et al. *Nature* **523,** 189–192 (2015).
3. Quimby, R. M. et al. *Nature* **474,** 487–489 (2011).
4. Aasi, J. et al. *Astrophys. J. Suppl. Ser.* **211,** 7(2014).
5. Levan, A. J. et al. *Astrophys. J.* **781,** 13 (2014).
6. Burrows, D. N. et al. *Nature* **476,** 421–424 (2011).
7. Gendre, B. et al. *Astrophys. J.* **766,** 30 (2013).
8. Stratta, G. et al. *Astrophys. J.* **779,** 66 (2013).
9. Inserra, C. et al. *Astrophys. J.* **770,** 128 (2013).
10. Kasen, D. & Bildsten, L. *Astrophys. J.* **717,** 245 (2010).
11. Woosley, S. E. *Astrophys. J.* **719,** L204 (2010).
12. Nicholl, M. et al. *Mon. Not. R. Astron. Soc.* **444,** 2096–2113 (2014).
13. Lunnan, R. et al. *Astrophys. J.* **787,** 138 (2014).
14. Chen, T.-W. et al. *Astrophys. J.* **763,** L28 (2013).

PROTISTOLOGY

How to build a microbial eye

Dissection of the subcellular eye of microorganisms called warnowiid dinoflagellates reveals that this structure is composed of elements of two cellular organelles — the plastid and the mitochondrion. SEE LETTER P.204

THOMAS A. RICHARDS & SUELY L. GOMES

The ancient Greek physician Galen described the key anatomical features of the eye[1], including the retina, lens, cornea and iris. Yet arguably the first true understanding of how the vertebrate eye works came in the early seventeenth century, with mathematician Johannes Kepler's demonstration that vision occurs as an image projected on to the surface of the retina[2]. As such, an eye can be defined as a cornea and/or a lens that forms an aperture allowing light arising from a specific direction to pass on to a sensory surface that processes this signal into a chemical message. But animals were not the only organisms to evolve such systems — analogous structures and biochemical responses exist in cells of several eukaryotic microorganisms (cells that package most of their DNA in a nucleus), allowing these microbes to move in response to light[3]. On page 204 of this issue, Gavelis *et al.*[4] describe the subcellular features that make up the eye-like structures of warnowiid dinoflagellates, which in anatomical terms are remarkably similar to vertebrate eyes.

Warnowiid dinoflagellates are unicellular plankton that have not been cultured in the laboratory, but that are known to possess a remarkably complex eye-like structure, called the ocelloid. Ocelloids consist of distinct components similar to key parts of vertebrate 'camera-type' eyes: a cornea, a lens (called a hyalosome) and a pigmented cup or retina-like structure. Gavelis *et al.* studied warnowiids isolated from marine waters in Japan and Canada, and demonstrate that the anatomy of ocelloids is built from reconfigured plastids and mitochondria (Fig. 1a). These are subcellular compartments seen in many eukaryotic groups that formed in the distant past through the intracellular incorporation of symbiotic bacteria; these organelles usually contain their own genomes and typically function in energy transformation.

Specifically, Gavelis and colleagues show that the retinal body of ocelloids arises from a membrane network derived from plastids, and that multiple mitochondria form a cornea-like surface across a lens structure. To test these microscopy-based observations, the authors microdissected the warnowiid
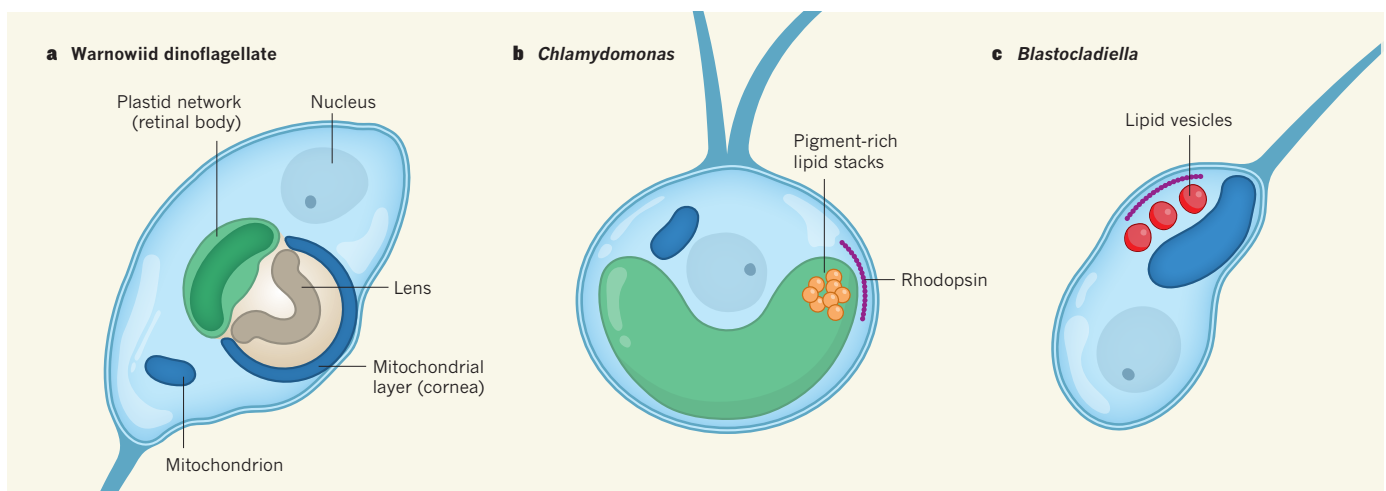
Figure 1 | Eyes across the tree of life. a, The eye-like ocelloids found in unicellular organisms known as warnowiid dinoflagellates have a 'camera-like' complexity that resembles that of animal eyes. Gavelis *et al.*[4] show that two of these components in warnowiids have arisen through the reconfiguration of membrane-bound organelles that are usually used for cellular energy transformation: the cornea is formed from a layer of mitochondria and the retinal body is derived from a network of plastids.

b, c, Microorganisms from other branches of the tree of life also contain eye-like structures, although these are anatomically simpler. **b,** The eyespots of *Chlamydomonas* algae comprise stacks of pigment-rich lipid molecules, located inside the cell's plastid, which shades light from one side of light-sensitive rhodopsin proteins. **c,** The eyespots of *Blastocladiella* fungi are lipid-filled vesicles close to the cell's main mitochondrion that are overlaid with rhodopsin proteins.

retinal body and sequenced its DNA, which contained a much higher proportion of DNA of plastid origin than equivalent samples from the whole cell.

Although ocelloids are exceptionally complex, warnowiids are not the only microbial cells with eye-like subcellular structures. A diversity of eukaryotic microorganisms perceive light using different kinds of eyespots. One such structure is the eyespot of the green alga *Chlamydomonas reinhardtii* (Fig. 1b), a unicellular relative of land plants. This eyespot is located at the edge of the alga's plastid and is made up of lipid globules, rich in orange carotenoid pigments, that are stacked in compartments inside the plastid envelope. As such, this globule layer is thought to provide directionality and contrast by shielding and reflecting light from one side of the organism on to two light-sensitive proteins called type 1 rhodopsins that localize with this eyespot[5-7]. These two proteins have intrinsic light-gated cation-channel activity (and are therefore named channelrhodopsins) and have been demonstrated to act as photoreceptors that trigger movement in response to light[5-7].

Cryptophyte algae such as *Guillardia theta* also build eyespot structures that are located in plastids[8], and movement of these cells in response to light is mediated by the function of at least two type 1 rhodopsin proteins[9], similar to *Chlamydomonas*. The alga *Euglena gracilis* also has an orange-red eyespot, although, in contrast to the previous examples, this structure is associated with the base of the flagellum[3], the cells' swimming propeller. The photoreceptor in *Euglena* has been identified as a photoactivated adenylyl cyclase[10] protein.

In yet another branch of the tree of life are the eyespot-like structures of the swimming spores of Blastocladiomycota fungi (Fig. 1c). These structures are lipid-filled vesicles called side-body complexes that are located close to the large mitochondrion of these fungal cells[11]. The side-body complex is overlaid with type 1 rhodopsin proteins. In *Blastocladiella emersonii*, the type 1 rhodopsin photosensor contains a guanylyl cyclase domain, which allows the protein to control the production of cyclic GMP (ref. 12), a key chemical messenger in vertebrate vision. Recent work[13] on warnowiid ocelloids has also suggested that messenger RNA encoding a type 1 rhodopsin is associated with the retinal body.

These examples demonstrate the wealth of subcellular structures and associated light-receptor proteins across diverse microbial groups. Indeed, all of these examples represent distinct evolutionary branches in separate major groups of eukaryotes[3]. Even the plastid-associated eyespots are unlikely to be the product of direct vertical evolution, because the *Chlamydomonas* plastid is derived from a primary endosymbiosis and assimilation of a cyanobacterium, whereas the *Guillardia* plastid is derived from a secondary endosymbiosis

in which the plastid was acquired 'second-hand' by intracellular incorporation of a red alga[14]. Using gene sequences recovered from the warnowiid retinal body, Gavelis *et al.* investigated the ancestry of this organelle by building phylogenetic trees for the plastid-derived genes. Their analysis demonstrated that this modified plastid is also of secondary endosymbiotic origin from a red alga.

Although derived independently, there are common themes in the evolution of these eye-like structures. Many of them involve the reconfiguration of cellular membrane systems to produce an opaque body proximal to a sensory surface, a surface that in four of the five examples probably involves type 1 rhodopsins. Given the evolutionary derivation of these systems, this represents a complex case of convergent evolution, in which photo-responsive subcellular systems are built up separately from similar components to achieve similar functions. The ocelloid example is striking because it demonstrates a peak in subcellular complexity achieved through repurposing

The ocelloid demonstrates a peak in subcellular complexity achieved through repurposing multiple components.

multiple components. Collectively, these findings show that evolution has stumbled on similar solutions to perceiving light time and time again. ∎

Thomas A. Richards *is in the Department of Biosciences, University of Exeter, Exeter EX4 4QD, UK.* **Suely L. Gomes** *is in the Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo 05508-000, Brazil. e-mails: t.a.richards@exeter.ac.uk; sulgomes@iq.usp.br*

1. Galen. *Galen on the Usefulness of the Parts of the Body* (Cornell Univ. Press, 1968).
2. Kepler, J. Astronomiae Pars Optica Manuscript presented to Rudolf II, Holy Roman Emperor (1604).
3. Jékely, G. *Phil. Trans. R. Soc. B* **364**, 2795–2808 (2009).
4. Gavelis, G. S. *et al. Nature* **523**, 204–207 (2015).
5. Schmidt, M. *et al. Plant Cell* **18**, 1908–1930 (2006).
6. Sineshchekov, O. A., Jung, K.-H. & Spudich, J. L. *Proc. Natl Acad. Sci. USA* **99**, 8689–8694 (2002).
7. Berthold, P. *et al. Plant Cell* **20**, 1665–1677 (2008).
8. Lucas, I. A. N. *Br. Phycol. J.* **17**, 13–19 (1982).
9. Sineshchekov, O. A. *et al. Biophys. J.* **89**, 4310–4319 (2005).
10. Ntefidou, M., Iseki, M., Watanabe, M., Lebert, M. & Häder, D.-P. *Plant Phys.* **133**, 1517–1521 (2003).
11. Lovett, J. S. *Bacteriol. Rev.* **39**, 345–404 (1975).
12. Avelar, G. M. *et al. Curr. Biol.* **24**, 1234–1240 (2014).
13. Hayakawa, S. *et al. PLoS ONE* **10**, e0118415 (2015).
14. Archibald, J. M. *Curr. Biol.* **19**, R81–R88 (2009).

MYELOID DISEASE

Another action of a thalidomide derivative

Lenalidomide effectively treats a blood disorder caused by the 5q chromosomal deletion. A study shows that the drug binds to its target, CRBN, to promote the breakdown of an enzyme encoded by a gene in the 5q region. SEE ARTICLE P.183

TAKUMI ITO & HIROSHI HANDA

Around 60 years ago, thalidomide was developed as a sedative and sold in more than 40 countries. But the drug was soon banned because of its association with serious developmental defects, such as limb deformities, in children whose mothers had taken it while pregnant. Now, thalidomide is being re-evaluated and is recognized as an effective treatment for myeloma, a cancer of plasma cells of the immune system. Moreover, derivatives of thalidomide have been developed; these compounds, which include lenalidomide and pomalidomide, make up a class of immunomodulatory drug termed IMiDs[1]. As well as being effective against myeloma, lenalidomide can treat[2] a type of myelodysplastic syndrome (MDS) — a disorder

of blood stem cells (haematopoietic cells) — that is caused by a deletion of the long arm of chromosome 5. In this issue, Krönke *et al.*[3] (page 183) provide a model of lenalidomide action in the context of this mutation.

The protein CRBN was identified as a direct target of thalidomide through affinity-bead technology[4]. CRBN functions as a substrate-recognition component of an E3 ubiquitin ligase enzyme complex that catalyses the conjugation of ubiquitin molecules to specific substrate proteins, thereby marking the proteins for degradation. CRBN is also bound by lenalidomide and pomalidomide[5,6] and is now regarded as a primary target of IMiDs — this binding is required for both the damaging and the therapeutic effects of the drugs. Previous research[7-9] showed that lenalidomide and pomalidomide promote the degradation
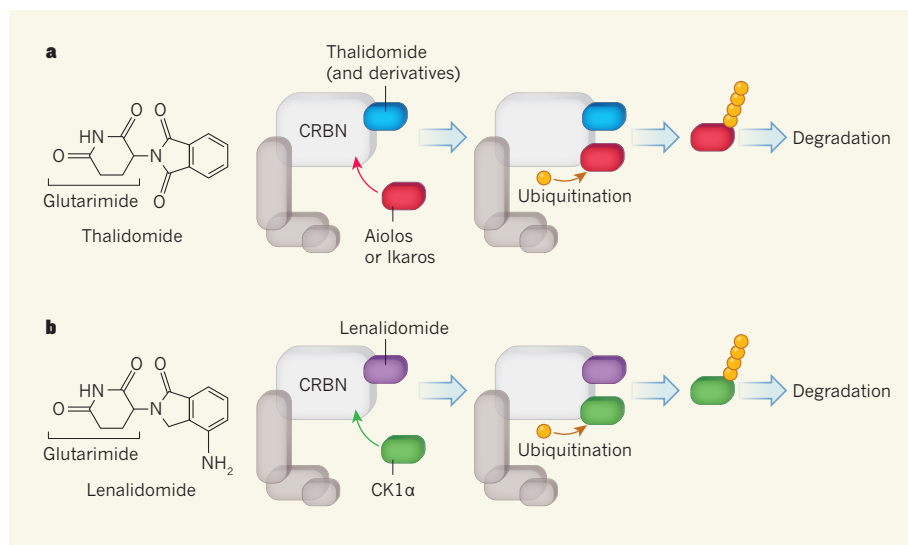
Figure 1 | Thalidomide and its derivatives confer the substrate specificity of CRBN. a, The immunomodulatory drugs thalidomide, lenalidomide, pomalidomide and CC-122 all bind to the protein CRBN, which is a substrate-recognition subunit of an E3 ubiquitin ligase enzyme complex. Binding of the drugs to CRBN induces the enzyme complex to attach ubiquitin molecules to the transcription factors Aiolos and Ikaros, which marks them for degradation[7–9]. This process explains the efficacy of these drugs for myelomas — cancers that arise from dysregulated proliferation of plasma cells, for which Aiolos and Ikaros are survival factors. **b**, Krönke et al. show that binding of CRBN by lenalidomide, but not the other related drugs, also induces ubiquitination and degradation of the regulatory enzyme CK1α. This activity underlies the drug's effective treatment of myelodysplastic syndrome (MDS), which is caused by the 5q chromosomal deletion and results in the loss of one copy of the gene encoding CK1α. The glutarimide moiety common to these drugs is marked.

of the transcription factors Ikaros (IKZF1) and Aiolos (IKZF3) by modulating the activity of a CRBN–ubiquitin ligase complex, and that this process underlies the drugs' efficacy against myeloma. However, it has been unclear whether all IMiDs confer the same substrate specificity on CRBN.

The deletion of the long arm of chromosome 5, called del(5q), is seen in some people with MDS, and leads to hyperproliferation of haematopoietic cells and their ineffective differentiation. Lenalidomide is known to selectively induce apoptotic cell death in cells with del(5q). The deletion means that people affected have only one copy of the genes located in that chromosomal region. Krönke et al. suggest that this haploinsufficiency might explain the efficacy of lenalidomide in this disease.

By studying the effect of lenalidomide treatment on protein ubiquitination and abundance of myeloid blood cells, the authors identify the enzyme casein kinase 1α (CK1α) as a target of ubiquitin-mediated degradation in the presence of the drug. Deletion of the *CRBN* gene using CRISPR/Cas9 genome-editing technology abolished this degradation, suggesting that this effect is crucially dependent on CRBN. The gene that encodes CK1α, *CSNK1A1*, is on the long arm of chromosome 5, so it seems that the result of this degradation is to compound the already lower than normal levels of this enzyme that result from the deletion.

CK1α regulates the activity of multiple proteins. For example, it negatively regulates p53, a tumour-suppressor protein. The CK1α

inhibitor D4476 has been shown to activate p53 and induce apoptosis in cells with only one copy of *CSNK1A1* (ref. 10). Krönke and colleagues demonstrate that CK1α depletion sensitizes normal human haematopoietic cells to lenalidomide. They also confirm that overexpression of CK1α confers lenalidomide resistance on cells with the del(5q) mutation. By contrast, overexpression of Ikaros did not suppress lenalidomide-mediated therapeutic effects on del(5q) cells.

Rodents have been shown to be resistant to IMiDs[1,6], and the authors found that lenalidomide did not decrease CK1α levels in normal mouse cells. They demonstrate that a single amino-acid difference between the human and mouse forms of CRBN is responsible for this different response to the drug, and that mouse cells expressing mouse CRBN with the substituted human amino acid were subject to lenalidomide-dependent CK1α degradation. The authors then generated 'humanized' mouse haematopoietic cells, which expressed the modified CRBN protein and had only a single copy of *CSNK1A1*. They show that lenalidomide treatment induced increased apoptotic death of the cells. Moreover, the increase in apoptosis in these cells was countered when p53 levels in the cells were reduced, which fits well with the previous report of p53 involvement in this pathway[10]. This development represents substantial progress in the field of IMiDs and thalidomide research — at last, through genetic modification of mouse CRBN, investigators can

use mice to study these drugs.

Krönke et al. then compared the effect of lenalidomide on a human MDS del(5q) cell line with that of other IMiDs, including CC-122, a compound currently undergoing phase I clinical trials for blood cancers and solid tumours. All IMiDs that were examined decreased levels of Aiolos and Ikaros, and the efficacy of CC-122 was stronger than that of lenalidomide. But only lenalidomide degraded CK1α (Fig. 1). Interestingly, high concentrations of CC-122 suppressed lenalidomide-induced CK1α degradation, which suggests that lenalidomide competes with CC-122 for binding to CRBN in cells and that lenalidomide confers a distinct substrate specificity on CRBN.

Accumulating evidence[3,7–9] indicates that the substrate recognition of CRBN is altered in response to each ligand that binds it. X-ray crystal structures of CRBN bound to IMiDs have revealed that a common glutarimide moiety in these compounds (Fig. 1) is sufficient for this binding to occur[5,6]. It is possible that the remaining structure of each ligand might be important for determining the enzyme's substrate specificity. Recently, uridine, a nucleotide base, was found to bind to the same ligand-binding pocket of CRBN[11], and it is likely that CRBN also has cellular ligands that alter its substrate specificity.

In plants, the hormone auxin functions as a molecular glue to attach the hormone's target TIR1, a substrate-recognition component of an SCF ubiquitin ligase, to its substrate AUX/IAA proteins[12]. It is unclear whether ligands such as IMiDs also function to glue CRBN to its substrate or whether they act through another mechanism. Elucidation of the 3D structure of the CRBN–ligand–substrate complex will provide a deeper understanding of the substrate-recognition process and contribute to the development of potent and clinically effective drugs. ∎

Takumi Ito *and* Hiroshi Handa *are in the Department of Nanoparticle Translational Research, Tokyo Medical University, Shinjuku-ku, Tokyo, 160-8402, Japan.*
e-mail: hhanda@tokyo-med.ac.jp

1. Ito, T., Ando, H. & Handa, H. *Cell. Mol. Life Sci.* **68**, 1569–1579 (2011).
2. List, A. *et al. N. Engl. J. Med.* **355**, 1456–1465 (2006).
3. Krönke, J. *et al. Nature* **523**, 183–188 (2015).
4. Ito, T. *et al. Science* **327**, 1345–1350 (2010).
5. Fischer, E. S. *et al. Nature* **512**, 49–53 (2014).
6. Chamberlain, P. P. *et al. Nature Struct. Mol. Biol.* **21**, 803–809 (2014).
7. Krönke, J. *et al. Science* **343**, 301–305 (2014).
8. Lu, G. *et al. Science* **343**, 305–309 (2013).
9. Gandhi, A. K. *et al. Br. J. Haematol.* **164**, 811–821 (2014).
10. Järås, M. *et al. J. Exp. Med.* **211**, 605–612 (2014).
11. Hartmann, M. D. *et al. J. Struct. Biol.* **188**, 225–232 (2015).
12. Tan, X. *et al. Nature* **446**, 640–645 (2007).

This article was published online on 1 July 2015.

REVIEW

Feedback in low–mass galaxies in the early Universe

Dawn K. Erb[1]

The formation, evolution and death of massive stars release large quantities of energy and momentum into the gas surrounding the sites of star formation. This process, generically termed 'feedback', inhibits further star formation either by removing gas from the galaxy, or by heating it to temperatures that are too high to form new stars. Observations reveal feedback in the form of galactic–scale outflows of gas in galaxies with high rates of star formation, especially in the early Universe. Feedback in faint, low–mass galaxies probably facilitated the escape of ionizing radiation from galaxies when the Universe was about 500 million years old, so that the hydrogen between galaxies changed from neutral to ionized—the last major phase transition in the Universe.

Although feedback from star formation is a complex phenomenon involving many processes over a wide range of physical scales, its most basic consequences are not difficult to understand: stars form out of gas, and therefore any process which either removes gas from a galaxy or prevents gas from condensing into new stars will have an important effect on the subsequent evolution of that galaxy. The topic of galactic-scale outflows in galaxies has been of interest to astronomers for half a century at least[1], but over the past two decades, the explosion of new facilities enabling large surveys of galaxies in the early Universe and the increasing sophistication of simulations of galaxy evolution have led to widespread recognition that outflows in galaxies at early times are both ubiquitous and essential: evidence for outflows is seen in nearly all star-forming galaxies at high redshifts[2–7], and without feedback, gas in galaxies would cool and form too many stars, resulting in much higher stellar masses than we observe today.

As the sizes of galaxy surveys increase and as new, more sensitive instruments are developed, we are beginning to constrain the properties and relative abundance of faint, low-mass galaxies in the early Universe. It is an opportune time to emphasize the importance of this population of galaxies: a new generation of sensitive, multi-object near-infrared spectrographs is offering access to their rest-frame optical emission lines (key to constraining the physical conditions in their star-forming regions) for the first time, while a variety of studies are pointing out the substantial contribution of such faint, low-mass galaxies to both the overall star-formation density of the Universe and the reionization of the Universe, when ionizing radiation from the first generation of stars and galaxies reionized the hydrogen gas in the intergalactic medium, which had been neutral since protons and electrons first combined 375,000 years after the Big Bang.

Complex, multi-phase galactic outflows are likely to determine the baryon and heavy-element content of both galaxies and the intergalactic medium. Outflows are seen in galaxies with unusually high star-formation rates in the local Universe, but the rest-frame ultraviolet spectra of galaxies at high redshifts show that outflows are a generic feature of star-forming galaxies in the early Universe. However, we have few constraints on the properties of feedback in the faint, low-mass (here defined as galaxies with stellar masses $\lesssim 10^9 M_\odot$) end of the distant galaxy population. Such constraints are badly needed: since outflow properties in the local Universe are observed to scale with mass and star-formation rate, an increase in the dynamic range over which these properties are measured at high redshifts will improve our understanding of the physical mechanisms behind feedback. Faint, low-mass galaxies are now also recognized as the likely source of many of the ionizing photons responsible for reionization, the last major phase transition in the Universe, and it is likely that feedback processes aided the escape of these photons. Current constraints on feedback in low-mass galaxies come mostly from Lyman α (Lyα) emission, which can indicate the presence of outflows but is unlikely to map directly to outflow velocity. However, future facilities such as the James Webb Space Telescope and the 30-m-class ground-based telescopes will provide a much more detailed view of this key process in low-mass galaxies.

An overview of galactic outflows

Both locally and in the distant Universe, galaxies with intense star formation concentrated in a small volume exhibit dramatic outflows of gas. The prototypical example of such a galaxy is the local starburst M82, shown in Fig. 1. M82 is a nearby disk galaxy with a powerful central starburst, thought to be triggered by interaction with its companion galaxy M81[9,10]. The nearly edge-on orientation of M82's disk reveals an extended, bi-conical outflow emerging perpendicular to the disk of the galaxy and emanating from its central starburst. Detailed, multi-wavelength observations of this outflow reveal its complex, multi-phase nature: hard X-ray emission (X-rays with energies $\gtrsim 10$ keV) traces hot gas with a temperature of 30–80 million K and terminal velocity of about 2,000 km s^{-1}, well above M82's escape velocity of about 460 km s^{-1} (ref. 11); softer X-rays and Hα emission from ionized gas indicate velocities of 600–800 km s^{-1} and suggest the presence of shocks[11,12]; and observations of CO emission indicate that the outflow also contains at least $3 \times 10^8 M_\odot$ of molecular gas, reaching a maximum velocity of about 230 km s^{-1} (ref. 13). As in other such galaxies, the mass of gas in the outflow is estimated to be comparable to the mass of gas being formed into stars[14]. M82 is thus the archetypal example of a galaxy in which the energy and momentum of star formation is heating gas to high temperatures and propelling at least some of it to velocities sufficient for the gas to escape the galaxy altogether. Since stars form out of gas, such outflows must have important consequences for the regulation of star formation in galaxies.

Although starburst galaxies such as M82 are relatively rare in the local Universe, the existence of galactic outflows in most star-forming galaxies at redshifts $z \approx 2$ and higher is now well-established[2,3,5,6]. It is also

[1]Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin Milwaukee, 3135 North Maryland Avenue, Milwaukee, Wisconsin 53211, USA.

Figure 1 | The local starburst galaxy M82 observed with the Hubble Space Telescope. Outflowing gas, seen in emission from ionized gas shown in red, forms a bi-conical structure centred on the intense starburst in the centre of the galactic disk. Image from NASA, ESA and The Hubble Heritage Team (STScI/AURA); http://hubblesite.org/newscenter/archive/releases/2006/14/image/a/.

Figure 2 | The ratio of stellar to halo masses as a function of halo mass in the local Universe. Light and dark shaded areas show 1σ and 2σ bounds respectively. The stellar to halo mass ratio peaks at halo mass $M_{halo} \approx 10^{12} M_\odot$, and declines at both lower and higher masses. The peak and slope depend somewhat on the redshift at which the relationship is measured, but the general form holds at all redshifts studied, and indicates that the efficiency of star formation decreases at both low and high masses. Image adapted from figure 4 of ref. 34 (IOP Publishing, American Astronomical Society).

now clear that the star-formation-rate density of the Universe peaks at $z \approx 2$–3 (ref. 15); star-formation feedback is therefore an essential part of the evolution of galaxies during the peak epoch of star formation, with a variety of important consequences for the properties of galaxies and the intergalactic medium. These consequences include the shape of the relationship between the heavy-element content of galaxies and their mass; in both local and high-redshift galaxies, more massive galaxies are more chemically enriched[16–18]. Elements heavier than helium (termed 'metals' by astronomers) are products of star formation and are released into the interstellar medium as stars evolve and die, enriching the gas and the next generation of stars. However, galactic outflows may drive some of these metals out of the galaxy instead, decreasing the overall metallicity and substantially modulating the relationship between metallicity and stellar mass[19–22].

Metals thus driven out of galaxies are observed in the gas between galaxies. Most of the baryons in the Universe lie not in galaxies but in the intergalactic medium[23], and observations of the gas between and around galaxies via absorption line systems in the spectra of quasars have long shown that some of the gas is metal-enhanced[24–26]. Individual metal absorption line systems can be associated with galaxies near the line of sight to the quasar[27], while large, joint surveys of quasars and foreground galaxies have enabled the statistical study of metal absorption lines arising from gas near galaxies[24,25,28,29]. Such studies have found that metals are enhanced near galaxies to distances from each galaxy of about 180 proper kiloparsecs (kpc) at $z \approx 2.4$ (ref. 30), while in the local Universe, observations probing the gas around galaxies to distances of 160 kpc have found that cool, metal-enriched gas accounts for as much as 25%–45% of the baryons in the halo of a typical $L^\star$ galaxy[31]. While the timescales over which metals are deposited into the intergalactic medium and the mixing of metals in the gas are still uncertain[32,33], it is becoming increasingly clear that galactic outflows are key to the regulation of the baryonic content of both galaxies and the gas around them.

A further issue that must be explained by any successful model of the formation and evolution of galaxies is the fact that the stellar fraction of galaxies is not constant[34,35]: as shown in Fig. 2, the ratio of galaxy stellar mass to the mass of the host dark-matter halo reaches a maximum at halo mass $M_{halo} \approx 10^{12} M_\odot$ (with some dependence on redshift), and declines at both higher and lower masses. In other words, both high- and low-mass galaxies are less efficient at forming stars. Feedback from active galactic nuclei has typically been invoked to explain the decrease in efficiency at high masses[36,37], although recent observations indicate

that fast outflows in some massive galaxies may be powered by starbursts rather than by active galactic nuclei[38,39]. At lower masses, the decreasing efficiency of star formation is likely to be due to processes associated with the formation and evolution of massive stars.

Galactic winds are driven by the energy and momentum imparted to gas by massive stars, but the relative importance of various feedback processes is still a subject of considerable study. Early work focused on the thermal energy provided by supernovae, which may drive a fast wind out of the starburst nucleus and into the galactic halo or beyond[40,41]. Many more recent models have focused on the role of momentum in driving galactic outflows; unlike the thermal energy injected by supernovae, momentum cannot be radiated away, and simple prescriptions for momentum-driven winds indicate mass outflow rates comparable to the star-formation rate and an inverse scaling of the wind efficiency (the mass outflow rate relative to the star-formation rate) with galaxy circular velocity, as suggested by observations of galactic outflows in the local Universe[42–44].

Modern models of galaxy formation and evolution emphasize the importance of feedback, but its implementation in simulations is difficult, primarily because of the range of scales involved (from cosmological scales to the scale of individual stars or at least giant molecular clouds) and the large number of complex and poorly understood physical processes. As a result, models rely on 'sub-grid' prescriptions for small-scale processes that are not resolved by the simulation. These models can reproduce observations and have demonstrated the effects of feedback on the evolution of galaxies and the intergalactic medium, but the results of these simulations may be dependent on the feedback prescriptions and how they are implemented[45,46].

Numerical simulations of galaxy formation are only now beginning to model feedback by directly modelling the physical processes involved. At present such simulations incorporate radiation pressure from starlight, energy and momentum injection from supernovae, stellar winds, and photoionization, and photoelectric heating; the results show that different feedback processes may interact in complex and nonlinear ways, and that different processes may dominate in different environments[47–49]. For example, radiation pressure probably dominates in massive, dusty and gas-rich starbursts, in which thermal energy is quickly radiated away, whereas the low gas densities of dwarf starbursts result in slow cooling times and hot winds driven by supernova heating[47]. In all cases, outflowing gas occupies a broad distribution in temperature and velocity, as seen in observations[14,50].

Observations of outflows at high redshifts

As exemplified by M82, the archetypal starburst described above, detailed observations of galactic outflows in the local Universe reveal that they are a complex, multi-phase phenomenon, with outflows observed in hot gas traced by X-rays[51,52], ionized gas seen via optical emission lines[53], neutral gas probed by low ionization absorption lines[14,50,54], and molecular gas observed at radio wavelengths[13,55]. These observations have revealed that the properties of outflows vary with galaxy mass, star-formation rate, and the surface density of star forma-tion: more massive galaxies with higher star-formation rates and higher star-formation-rate densities tend to drive faster outflows[14,50,56]. An observational determination of the scalings of outflow properties with galaxy properties is key to identifying the underlying mechanisms driving the outflow[42,47], but this determination is complicated by the fact that correlations appear to be relatively weak and may flatten at higher masses and star-formation rates. For example, studies find the wind velocity is proportional to star-formation rate (SFR) as follows: $v \propto \text{SFR}^{\alpha}$, with α ranging from 0.1 to 0.35. The result is that a wide dynamic range in galaxy properties, extending to dwarf galaxies, is required to detect trends robustly[4,14,50]. Such measurements are possible in the local Universe, where the masses of the galaxies studied range from about $10^7 M_{\odot}$ to $10^{11} M_{\odot}$ and their star-formation rates range from about $10^{-2} M_{\odot}$ per year to $10^3 M_{\odot}$ per year[50,56].

At redshifts $z > 1$, observations are much more limited, owing to both the faintness of galaxies at these distances and the redshifting of key diagnostics into the near-infrared; the result is that high-redshift sam-ples used to study outflows contain few galaxies with masses $< 10^9 M_{\odot}$ or star-formation rates less than a few $M_{\odot}$ per year. While outflows in relatively massive galaxies have been detected via observations of ionized[7] and molecular gas[57], the vast majority of our observational knowledge of feedback in galaxies at high redshift (here referring to redshifts $1 \lesssim z \lesssim 4$, at which most observations of feedback in distant galaxies have been made) comes from observations of interstellar absorption and Lyα emission lines in the rest-frame ultraviolet spectra of these galaxies. A typical example of such a spectrum is shown in Fig. 3. Because this spectrum traces the rest-frame ultraviolet, the continuum light is produced by hot, massive young stars. The strong absorption features are resonance lines from interstellar gas, blended at this reso-lution to include both the interstellar medium of the galaxy and gas entrained in outflows. This spectrum shows Lyα in emission, but such

spectra may exhibit anything from strong emission to strong absorption, or a superposition of the two.

The first challenge that arises when studying galactic outflows from a spectrum such as this is the determination of the velocity zero point from which to measure outflow velocities; all of the strong features in this spectrum come from interstellar gas, which is not necessarily at rest with respect to the stars. Thus the systemic redshift, the redshift of the stars, must be measured from stellar features or from emission lines emanating from ionized gas surrounding the stars. Stellar features are typically too broad and weak to be measurable in spectra of distant galaxies, so observers usually measure systemic redshifts from the rest-frame optical emission lines of ionized gas, which are redshifted into the near-infrared for $z \gtrsim 1.5$. Observations in the near-infrared have historically been considerably more difficult than in the optical because of the higher sky background and the relative inefficiency of detectors, but this situation is now changing with the advent of new multi-object near-infrared spectrographs on large telescopes[58–60].

Once the systemic redshift can be determined, a characteristic pattern emerges, visible in the spectrum shown in Fig. 3: the interstellar absorp-tion lines are blueshifted with respect to the systemic velocity, while Lyα emission, if present, is redshifted. A simple schematic of the model giving rise to this pattern is shown in Fig. 4. The blueshifted interstellar absorption lines are due to the absorption of light from the stellar con-tinuum by foreground gas moving towards the observer in a galactic outflow (though there may also be absorption from non-outflowing gas in the interstellar medium of the galaxy), and the strength of the absorp-tion lines depends on the covering fraction of this outflowing gas. Lyα emission is produced in H II regions surrounding the massive stars and is then strongly modified by resonant scattering: Lyα photons are absorbed and re-emitted by neutral hydrogen in and around the galaxy, with the result that they are most likely to escape the galaxy in the direction of the observer when they are backscattered from gas on the far, receding side of the outflow and thereby acquire a frequency shift that allows them to pass through the gas in the bulk of the galaxy unimpeded. This frequency shift then results in the redshift of Lyα emission relative to the systemic velocity. Galaxies at $z \approx 2$–3 show typical absorption line blueshifts of about 200 km s^{-1}, while Lyα is typically redshifted by about 500 km s^{-1}; however, measurements of individual galaxies reveal substantial variation in both quantities[2,3,5].
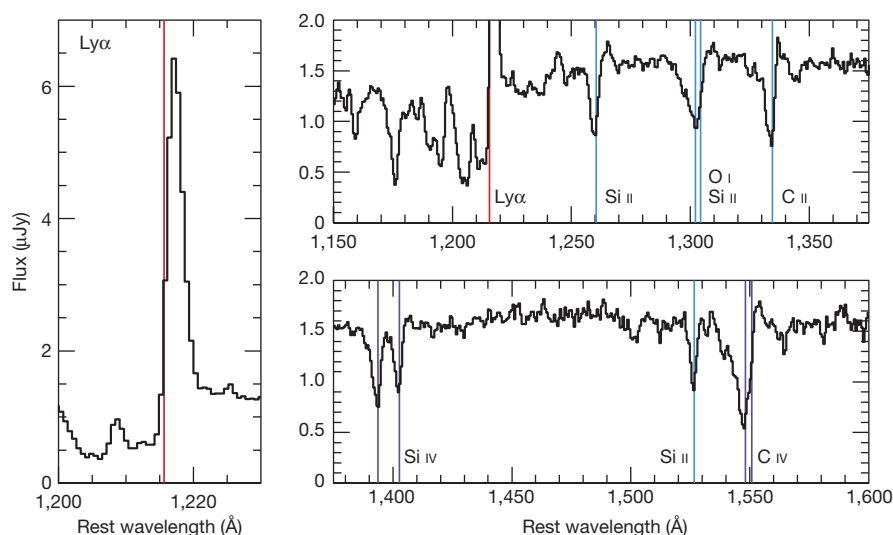
Figure 3 | A moderately low-resolution rest-frame ultraviolet spectrum of a typical star-forming galaxy at redshift $z \approx 2$–3. This spectrum, of a galaxy at $z = 2.33$ and shown in the rest frame, represents 23 h of integration with the Low Resolution Imaging Spectrometer (LRIS) on the 10-m Keck I telescope, and is therefore considerably higher in signal-to-noise ratio than most existing spectra of galaxies at comparable redshifts. The Lyα emission line is shown in the left panel, and the two right panels show continuum emission from hot, massive stars, on which is superimposed resonance absorption lines from interstellar gas. Absorption is present from both low- and high-ionization transitions, and the strongest features are marked and labelled. Red lines indicate Lyα emission, blue lines indicate low ionization absorption lines, and purple lines indicate high ionization absorption lines.
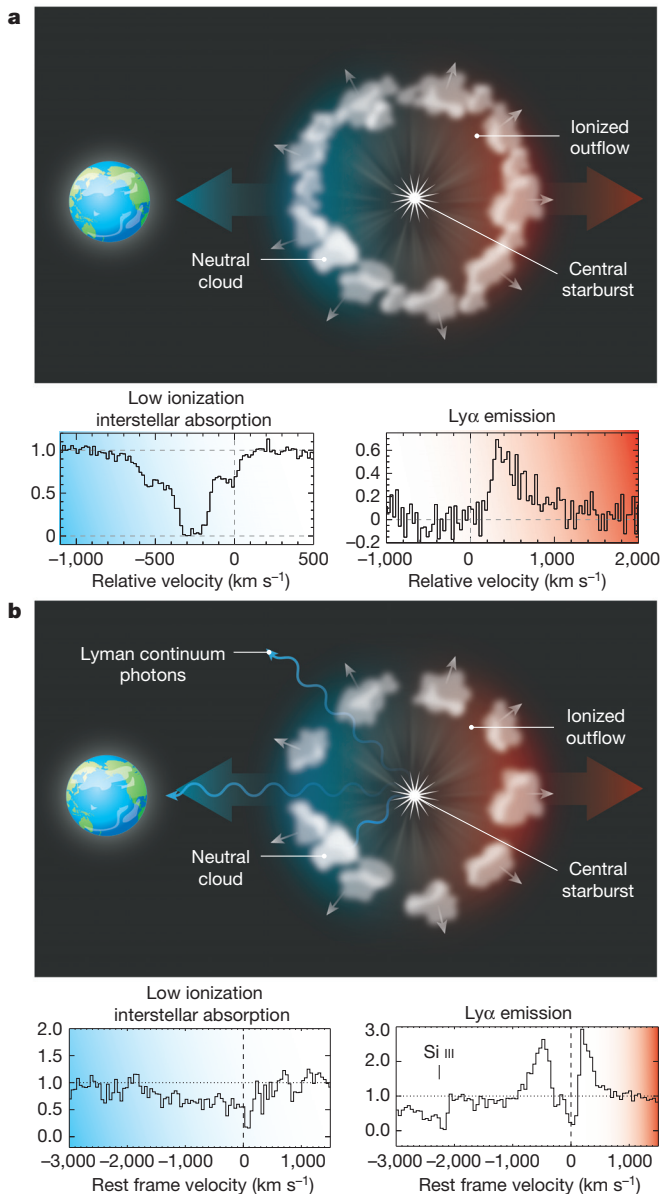
Figure 4 | Two schematic models of a spherical galactic outflow. a, A model in which the covering fraction of neutral hydrogen in the outflow is nearly complete. **b**, The covering fraction of neutral gas is lower, resulting in substantial residual intensity in the low ionization absorption lines, Lyα emission that is blueshifted as well as redshifted, and the potential escape of ionizing Lyman continuum photons. The spectra in **a** show low ionization absorption and Lyα emission from the $z = 2.7$ lensed galaxy MS1512-cB58[61], and the spectra in **b** show absorption and emission from the $z = 0.23$ galaxy J0921+4509[56,82]. Spectra in **a** courtesy of M. Pettini. Spectra in **b** adapted from figures 4 and 8 of ref. 56 (IOP Publishing, American Astronomical Society).

A further difficulty with observations of galactic outflows at high redshift is the low spectral resolution at which such observations are generally made, since these galaxies are typically too faint for observations at high resolution. Low resolution results in blending between absorption lines from outflowing and non-outflowing components of gas in the galaxy, with the result that the centroid of the interstellar absorption lines is a crude measurement of the outflow velocity at best; this centroid may be strongly influenced by the strength and width of absorption from gas at the redshift of the galaxy itself. This problem can be improved with the use of spectra of gravitationally lensed galaxies, in which magnification by a foreground galaxy or cluster of galaxies can result in the amplification of flux by a factor of 30 or more.

The brightness of these galaxies then allows for higher-resolution spectra in which the velocity structure of the interstellar gas can be studied in much more detail. An example of absorption lines in a gravitationally lensed galaxy is shown in Fig. 5. These lines show an absorption component at zero velocity, corresponding to the interstellar medium of the galaxy, a strong outflowing component centred at approximately -250 km s^{-1}, and a tail of outflowing gas with velocities extending to -750 km s^{-1} (ref. 61). More recent studies of additional lensed galaxies, and of high-signal-to-noise-ratio composite spectra of large samples of galaxies, indicate that such maximum velocities are typical, with winds extending to velocities of -800 km s^{-1} (refs 5, 62–64).

Examinations of the ultraviolet spectra of galaxies at $1.5 \lesssim z \lesssim 4$ have indicated that outflows are prevalent at these redshifts, and have provided estimates of their typical velocities, but many questions remain. One such question is whether or not the scalings of outflow properties observed in the local Universe also hold at high redshift. At $z \approx 1.4$, outflow velocity is observed to scale with star-formation rate with a comparable scaling to the local relation[4], but studies at higher redshifts have shown mixed or inconclusive results, possibly owing to the lack of dynamic range in the samples because of the inability to measure outflow velocities in very faint galaxies at high redshifts. Thus, the inclusion of faint objects in spectroscopic samples at high redshifts is key to understanding the properties of feedback in the early Universe.

The importance of low-mass galaxies

While observations of low-mass galaxies (here defined as galaxies with stellar masses $\lesssim 10^9 M_\odot$) are necessary to an understanding of how feedback operates in the distant Universe, these objects are also important in their own right. Measurements of the rest-frame ultraviolet luminosity function of galaxies indicate that, by $z > 0.75$, the faint-end slope is steeper than in the local Universe, and it remains steep and may even increase out to the highest redshifts at which it can be measured[65–68]. Studies of samples of galaxies lensed by massive clusters also indicate that this slope remains steep down to the faintest observable magnitudes[69,70]. These results indicate that faint, low-mass galaxies host a substantial fraction of the star formation in the high-redshift Universe, while also making it clear that the determination of the contribution of faint galaxies to the global density of star formation depends on assumptions regarding the stellar populations of these faint galaxies. The metallicities, dust properties and ages of these objects are not yet well characterized[65,69].

Faint galaxies are now also being recognized as the probable key to the reionization of the Universe. Ionizing photons from the first stars and galaxies reionized the intergalactic medium, and observations now constrain the epoch at which this occurred. Spectra of quasars at $z > 6$ reveal broad, total absorption at wavelengths just short of the Lyα emission line in the spectrum of the quasar itself (the Gunn–Peterson effect), indicating the presence of neutral hydrogen in the surrounding intergalactic medium and thus suggesting the completion of reionization at $z \approx 6$ (ref. 71). Observations of the cosmic microwave background also constrain the redshift of reionization, through the increased optical depth as cosmic microwave background photons scatter off newly free electrons. Measurements of this optical depth place the redshift of reionization at $z \approx 10$–11, assuming that it was a nearly instantaneous process[72,73]; it is more likely, however, that reionization proceeded more gradually, beginning at $z > 10$ and completing at $z \approx 6$–7 (ref. 74). The production and escape of sufficient numbers of ionizing photons remains a challenge for models of reionization, with current models suggesting that large numbers of faint galaxies are required[75,76].

The optical depth of the intergalactic medium precludes detection of ionizing photons at all redshifts $z \gtrsim 4$, but studies at $z \approx 3$ now suggest that the escape fraction may be higher in faint galaxies selected by Lyα emission than in brighter, continuum-selected Lyman break galaxies (galaxies identified via broad-band imaging in filters bracketing the drop in flux at the ionization edge of hydrogen)[77–79]. Although many uncertainties
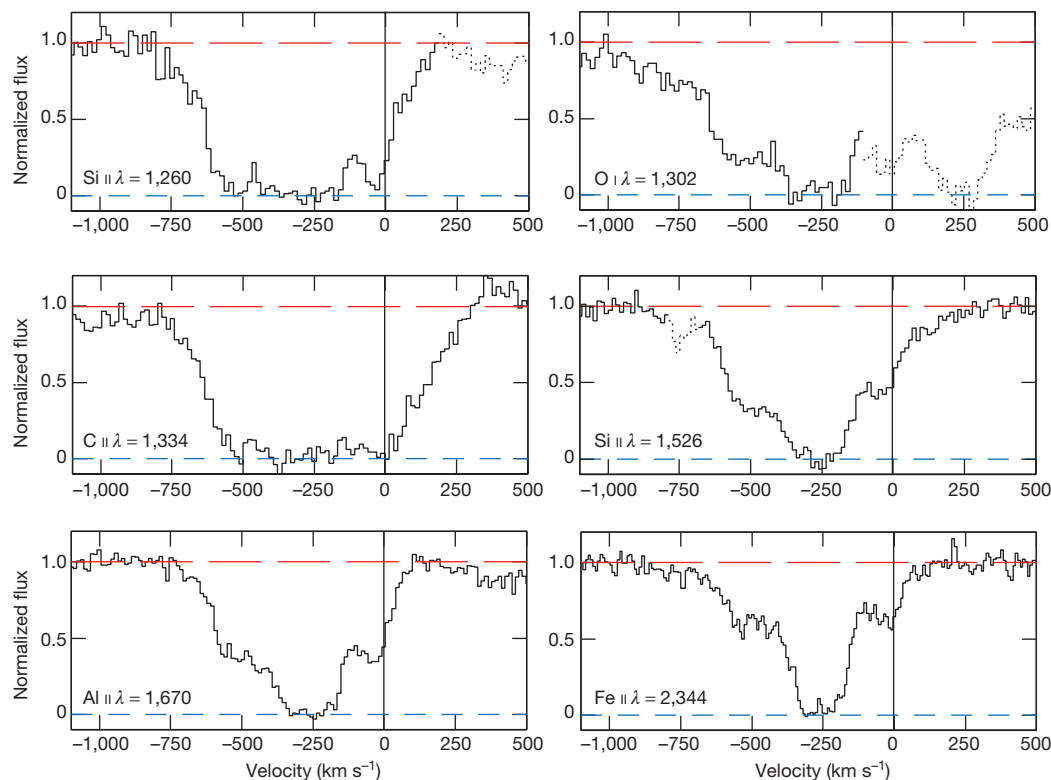
Figure 5 | Low ionization interstellar absorption line profiles of the $z = 2.73$ gravitationally lensed galaxy MS1512-cB58. The gas with highest optical depth has a velocity of about $250\,\mathrm{km\,s^{-1}}$, but absorption is seen over a wide range of velocities, up to $750\,\mathrm{km\,s^{-1}}$. Dotted lines show transitions other than the one labelled in each panel. Image adapted from figure 1 of ref. 61 (IOP Publishing, American Astronomical Society).

remain, observations and models are beginning to build a picture of galaxies that may leak ionizing radiation. Lyman continuum photons are likely to escape from galaxies through holes in a patchy distribution of neutral hydrogen[79], and galactic winds and photoionization from massive stars are expected to be key to clearing out such channels. Galaxies locally and at $z \approx 2$ have been observed to have weak but saturated low ionization interstellar absorption lines, suggesting partial coverage of neutral gas;[56,80] they are also characterized by noticeable Lyα emission emerging blueward of the systemic velocity (see further discussion below), again characteristic of a low covering fraction of neutral hydrogen. At high redshift, these signatures are associated with intense, low-metallicity starbursts with little dust[80,81], while in the local Universe, where observations with higher spatial and spectral resolution are possible, escaping Lyman continuum emission has been directly detected from a Lyman break analogue galaxy with intense star formation and partial coverage of neutral gas, as indicated by the depths of absorption lines[82]. The general picture arising from all of these observations is of a compact, possibly low-metallicity galaxy in which strong star formation in a small volume drives powerful feedback; the combination of strong winds and intense ionizing radiation then results in incomplete coverage of neutral hydrogen, allowing some of the ionizing photons to escape. Such a scenario is illustrated in Fig. 4b.

Galactic outflows and Lyα emission

As described above, the most direct probe of outflowing gas in distant galaxies is absorption line spectroscopy. Given sufficient spectral resolution and sufficiently high signal-to-noise ratio, such spectra provide a map of the covering fraction of absorbing gas as a function of velocity, for both low and higher ionization states. Spectra with lower resolution and/or lower signal-to-noise ratio have provided valuable results through the use of long integration times or the stacking of large numbers of spectra[3,6], but the technique fundamentally requires the spectroscopic detection of continuum emission with at least a moderate signal-to-noise ratio, making its application to very faint, distant galaxies extremely challenging with current technology.

A more immediately accessible but more difficult to interpret probe of gas in galaxies at high redshifts is provided by Lyα emission. Distant galaxies can be relatively easily selected at a particular redshift by taking deep images with a narrowband filter with a central wavelength corresponding to the wavelength of Lyα at the redshift of interest[83]. Galaxies at $z \approx 2$–3 selected in this way are fainter on average than galaxies selected in typical magnitude-limited surveys at the same redshifts, with typical stellar masses of $(3–10) \times 10^8 M_\odot$ and little dust extinction[84–87]. Once identified, these galaxies can be studied spectroscopically; although they are faint, the combination of measurements of their Lyα profiles and systemic redshifts from rest-frame optical emission lines (shifted into the near-infrared for $z > 1.5$, which has made it difficult to obtain large samples) yields valuable constraints on the presence of galactic outflows and the covering fraction and column density of neutral hydrogen[86,88–94].

As shown in Figs 3 and 4, asymmetric, redshifted Lyα emission is a signature of outflowing gas. However, mapping the velocity profile of the Lyα emission line to the velocity structure of the outflowing gas is challenging, because the Lyα profile is affected by many other factors in addition to the gas velocity. Radiative transfer models show that the strength and velocity offset of Lyα emission depend not only on the kinematics of the outflowing gas, but on its column density, covering fraction, dust content, the angle at which the galaxy is viewed, and the presence of non-outflowing neutral hydrogen at the systemic velocity[88,95–99]. Thus, without additional constraints on these parameters, tracing the velocity structure of outflowing gas via Lyα emission is extremely difficult.

Lyα emission nevertheless provides a valuable probe of the kinematics and physical conditions of gas in galaxies that are otherwise extremely difficult to study spectroscopically. Studies of the velocity offset of Lyα emission from the systemic velocity at $z \approx 2$–3 indicate that Lyα emission is typically redshifted, even in the faintest, lowest-mass galaxies in which it has been measured (Hubble Space Telescope F814W magnitudes $m_{AB} \approx 27$, dynamical masses $M_{dyn} < 10^8 M_\odot$), indicative of the presence of outflowing gas in most faint objects studied. These studies

also indicate that the velocity offset of Lyα emission increases with increasing rest-frame ultraviolet luminosity and nebular line velocity dispersion, implying that more-massive galaxies with higher star-formation rates have larger velocity offsets (see Fig. 6), and that lower-equivalent-width Lyα emission is also associated with larger velocity offsets[89,93,94].

The most obvious explanation for lower Lyα velocities in faint galaxies is that the Lyα emission is tracing a slower outflow. This is plausible; as described above, in the local Universe and to $z \approx 1.5$, the outflow velocity as traced by the centroids of absorption lines scales with both mass and star-formation rate[4,14,50,56], so a decrease in wind velocity would be unsurprising. However, this is not the only explanation for the observed trends. A further clue to the interpretation of Lyα emission in faint galaxies lies in the anti-correlation between Lyα velocity offset and equivalent width. Galaxies with stronger Lyα emission also have smaller Lyα velocity offsets, probably because an increase in the column density, covering fraction or velocity dispersion of neutral gas in the galaxy will both require that Lyα photons attain larger velocity shifts in order to escape the galaxy and increase the probability that they will be absorbed by dust during multiple scatterings or scattered beyond the spectroscopic slit[94]. Thus a decrease in the velocity offset of Lyα emission may be associated with the development of a stable, gaseous disk as galaxies grow[100].

Although Lyα emission from strongly star-forming galaxies is typically redshifted, some galaxies also show substantial Lyα emission emerging blueward of the systemic velocity[56,80,82,94]. Such blueshifted emission appears to be associated with intense, compact star formation, and appears in local galaxies with compact central sources driving high-velocity outflows[56] and at higher redshifts in galaxies with spectra indicative of low metallicity and high ionization parameters[80]. Recent observations of very faint, low-mass (approximately $(10^6 - 10^9)M_\odot$), lensed galaxies at $z \approx 2$ also indicate that a low metallicity and high ionization state are characteristic of low-mass galaxies, although the outflow properties of these lensed galaxies have not yet been constrained[81]. The implication (inferred also from weak but saturated low ionization absorption lines and stronger high ionization lines) is that much of the outflowing gas in such low-mass, low-metallicity galaxies is highly ionized, allowing Lyα emission to emerge with relatively little

scattering by neutral hydrogen[56,80]. An absence of neutral hydrogen is also required for the escape of Lyman continuum emission, implying that galaxies with a relatively high fraction of blueshifted Lyα emission may be good candidates for Lyman continuum emission[94]; indeed, Lyman continuum photons have recently been detected from one such local galaxy[82]. Low-mass galaxies with compact star formation resulting in a highly ionized outflow may thus be the best candidates for the source of the bulk of the photons responsible for the reionization of the Universe.

The next steps

Spectroscopic observations have now established the near ubiquity of gaseous outflows in star-forming galaxies at high redshifts, but many questions remain. There are currently very few constraints on the properties and prevalence of feedback in faint (optical $m_{AB} \gtrsim 25.5$), low-mass galaxies in the early Universe, and in part because such feedback may clear channels for the escape of the energetic photons needed to reionize the Universe, such constraints are badly needed. Direct determinations of outflow velocities in faint objects require absorption line spectroscopy; such determinations will come from long integrations, stacking very large samples, and observations of lensed galaxies. All of this work is in progress, and results will come in the next few years. Multi-object near-infrared spectrographs have already enabled systemic redshifts to be determined for larger samples of fainter galaxies than previously possible, and larger, deeper samples will constrain the physical conditions in these galaxies, relating the kinematics and covering fraction of outflowing neutral gas to the physical conditions in star-forming regions.

The next breakthroughs will be provided by future facilities. Upcoming 30-m-class telescopes (the Giant Magellan Telescope, the Thirty Meter Telescope, and the Extremely Large Telescope) will enable rest-frame ultraviolet absorption line spectroscopy of fainter galaxies at higher resolution and extending to higher redshifts; such higher-resolution observations will provide maps of the covering fraction of outflowing gas as a function of velocity. While these observations will still be limited by spectral resolution and signal-to-noise ratio, especially for fainter objects, they will undoubtedly extend our quantitative knowledge of galactic outflows at high redshift to include a much wider range of galaxy properties than is currently possible. The James Webb Space Telescope will enable near-infrared spectroscopy of galaxies at $z \gtrsim 3.5$ (the redshift at which most of the strong rest-frame optical emission lines shift beyond the ground-based atmospheric windows), allowing the determination of systemic redshifts and physical conditions in much more distant galaxies. In combination, these large ground- and space-based observatories will transform our view of the properties, prevalence, mechanisms and impact of feedback in the distant Universe.

Figure 6 | The velocity offset of Lyα emission as a function of ultraviolet luminosity and velocity dispersion, for $z \approx 2$–3 galaxies selected by strong Lyα emission. a, The velocity offset of Lyα emission from the systemic velocity versus rest-frame ultraviolet luminosity, shown in units of the solar luminosity on the bottom axis and in absolute ultraviolet magnitude on the top axis. The dashed vertical lines indicate the characteristic luminosity $L^\star$ of galaxies at $z \approx 2$ (red) and $z \approx 3$ (purple)[65]. b, The velocity offset of Lyα emission from the systemic velocity versus velocity dispersion, for the same sample of galaxies. The velocity dispersion is measured from the width of nebular emission lines, and is an indication of the depth of the gravitational potential well. Data in this figure are taken from figures 4 and 5 of ref. 94 (IOP Publishing, American Astronomical Society).

1. Burke, J. A. Mass flow from stellar systems—I. Radial flow from spherical systems. *Mon. Not. R. Astron. Soc.* **140,** 241–254 (1968).
2. Pettini, M. *et al.* The rest-frame optical spectra of Lyman break galaxies: star formation, extinction, abundances, and kinematics. *Astrophys. J.* **554,** 981–1000 (2001).
   This paper is the first study of rest-frame optical emission lines at $z \approx 3$ that indicated the ubiquity and importance of galactic outflows.
3. Shapley, A. E., Steidel, C. C., Pettini, M. & Adelberger, K. L. Rest-frame ultraviolet spectra of $z \approx 3$ Lyman break galaxies. *Astrophys. J.* **588,** 65–89 (2003).
   This paper is a comprehensive study of the rest-frame ultraviolet spectra of star-forming galaxies at $z \approx 3$, and demonstrates that the trends in the spectra arise from the properties of galactic outflows.
4. Weiner, B. J. *et al.* Ubiquitous outflows in DEEP2 spectra of star-forming galaxies at z = 1.4. *Astrophys. J.* **692,** 187–211 (2009).
5. Steidel, C. C. *et al.* The structure and kinematics of the circumgalactic medium from far-ultraviolet spectra of $z \approx 2$–3 galaxies. *Astrophys. J.* **717,** 289–322 (2010).
6. Jones, T., Stark, D. P. & Ellis, R. S. Keck spectroscopy of faint 3 < z < 7 Lyman break galaxies: III. The mean ultraviolet spectrum at $z \approx 4$. *Astrophys. J.* **751,** 51 (2012).
7. Newman, S. F. *et al.* The SINS/zC-SINF survey of z = 2 galaxy kinematics: outflow properties. *Astrophys. J.* **761,** 43 (2012).

8. White, S. D. M. & Rees, M. J. Core condensation in heavy halos—a two-stage theory for galaxy formation and clustering. *Mon. Not. R. Astron. Soc.* **183,** 341–358 (1978).

9. Telesco, C. M., Joy, M., Dietz, K., Decher, R. & Campins, H. Infrared mapping of M82—a starburst in an edge-on barred galaxy. *Astrophys. J.* **369,** 135–146 (1991).

10. Yun, M. S., Ho, P. T. P. & Lo, K. Y. H. I streamers around M82—tidally disrupted outer gas disk. *Astrophys. J.* **411,** L17–L20 (1993).

11. Strickland, D. K. & Heckman, T. M. Supernova feedback efficiency and mass loading in the starburst and galactic superwind exemplar M82. *Astrophys. J.* **697,** 2030–2056 (2009).

12. Lehnert, M. D., Heckman, T. M. & Weaver, K. A. Very extended X-ray and Hα emission in M82: implications for the superwind phenomenon. *Astrophys. J.* **523,** 575–584 (1999).

13. Walter, F., Weiss, A. & Scoville, N. Molecular gas in M82: resolving the outflow and streamers. *Astrophys. J.* **580,** L21–L25 (2002).

14. Rupke, D. S., Veilleux, S. & Sanders, D. B. Outflows in infrared-luminous starbursts at z < 0.5. II. Analysis and discussion. *Astrophys. J.* (Suppl.) **160,** 115–148 (2005).

15. Madau, P. & Dickinson, M. Cosmic star-formation history. *Annu. Rev. Astron. Astrophys.* **52,** 415–486 (2014).

16. Tremonti, C. A. *et al.* The origin of the mass-metallicity relation: insights from 53,000 star-forming galaxies in the Sloan Digital Sky Survey. *Astrophys. J.* **613,** 898–913 (2004).

17. Erb, D. K. *et al.* The mass-metallicity relation at z ≥ 2. *Astrophys. J.* **644,** 813–828 (2006).

18. Maiolino, R. *et al.* AMAZE. I. The evolution of the mass-metallicity relation at z > 3. *Astron. Astrophys.* **488,** 463–479 (2008).

19. Dalcanton, J. J. The metallicity of galaxy disks: infall versus outflow. *Astrophys. J.* **658,** 941–959 (2007).

20. Erb, D. K. A model for star formation, gas flows, and chemical evolution in galaxies at high redshifts. *Astrophys. J.* **674,** 151–156 (2008).

21. Finlator, K. & Davé, R. The origin of the galaxy mass-metallicity relation and implications for galactic outflows. *Mon. Not. R. Astron. Soc.* **385,** 2181–2204 (2008).

22. Davé, R., Finlator, K. & Oppenheimer, B. D. An analytic model for the evolution of the stellar, gas and metal content of galaxies. *Mon. Not. R. Astron. Soc.* **421,** 98–107 (2012).

23. Davé, R., Oppenheimer, B. D., Katz, N., Kollmeier, J. A. & Weinberg, D. H. The intergalactic medium over the last 10 billion years—I. Lyα absorption and physical conditions. *Mon. Not. R. Astron. Soc.* **408,** 2051–2070 (2010).

24. Prochaska, J. X., Weiner, B., Chen, H.-W., Mulchaey, J. & Cooksey, K. Probing the intergalactic medium/galaxy connection. V. On the origin of Lyα and O VI absorption at z < 0.2. *Astrophys. J.* **740,** 91 (2011).

25. Rudie, G. C. *et al.* The gaseous environment of high-z galaxies: precision measurements of neutral hydrogen in the circumgalactic medium of z ≈ 2–3 galaxies in the Keck Baryonic Structure Survey. *Astrophys. J.* **750,** 67 (2012).

26. Cowie, L. L., Songaila, A., Kim, T.-S. & Hu, E. M. The metallicity and internal structure of the Lyman-alpha forest clouds. *Astron. J.* **109,** 1522–1530 (1995).

27. Simcoe, R. A., Sargent, W. L. W., Rauch, M. & Becker, G. Observations of chemically enriched QSO absorbers near z ≈ 2.3 galaxies: galaxy formation feedback signatures in the intergalactic medium. *Astrophys. J.* **637,** 648–668 (2006).

28. Steidel, C. C., Dickinson, M. & Persson, S. E. Field galaxy evolution since Z approximately 1 from a sample of QSO absorption-selected galaxies. *Astrophys. J.* **437,** L75–L78 (1994).

29. Adelberger, K. L., Steidel, C. C., Shapley, A. E. & Pettini, M. Galaxies and intergalactic matter at redshift z ≈ 3: overview. *Astrophys. J.* **584,** 45 (2003).

30. Turner, M. L., Schaye, J., Steidel, C. C., Rudie, G. C. & Strom, A. L. Metal-line absorption around z ≈ 2.4 star-forming galaxies in the Keck Baryonic Structure Survey. *Mon. Not. R. Astron. Soc.* **445,** 794–822 (2014).

31. Werk, J. K. *et al.* The COS-Halos Survey: physical conditions and baryonic mass in the low-redshift circumgalactic medium. *Astrophys. J.* **792,** 8 (2014).

32. Simcoe, R. A. The carbon content of intergalactic gas at z = 4.25 and its evolution toward z = 2.4. *Astrophys. J.* **738,** 159 (2011).

33. Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A. & Weinberg, D. H. The intergalactic medium over the last 10 billion years—II. Metal-line absorption and physical conditions. *Mon. Not. R. Astron. Soc.* **420,** 829–859 (2012).

34. Moster, B. P. *et al.* Constraints on the relationship between stellar mass and halo mass at low and high redshift. *Astrophys. J.* **710,** 903–923 (2010).

35. Behroozi, P. S., Conroy, C. & Wechsler, R. H. A comprehensive analysis of uncertainties affecting the stellar mass-halo mass relation for 0 < z < 4. *Astrophys. J.* **717,** 379–403 (2010).

36. Hopkins, P. F. *et al.* A unified, merger-driven model of the origin of starbursts, quasars, the cosmic X-ray background, supermassive black holes, and galaxy spheroids. *Astrophys. J.* (Suppl.) **163,** 1–49 (2006).

37. Croton, D. J. *et al.* The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365,** 11–28 (2006).

38. Sell, P. H. *et al.* Massive compact galaxies with high-velocity outflows: morphological analysis and constraints on AGN activity. *Mon. Not. R. Astron. Soc.* **441,** 3417–3443 (2014).

39. Geach, J. E. *et al.* Stellar feedback as the origin of an extended molecular outflow in a starburst galaxy. *Nature* **516,** 68–70 (2014).

40. Chevalier, R. A. & Clegg, A. W. Wind from a starburst galaxy nucleus. *Nature* **317,** 44–45 (1985).

41. Veilleux, S., Cecil, G. & Bland-Hawthorn, J. Galactic winds. *Astron. Rev. Astron. Astrophys.* **43,** 769–826 (2005).

42. Murray, N., Quataert, E. & Thompson, T. A. On the maximum luminosity of galaxies and their central black holes: feedback from momentum-driven winds. *Astrophys. J.* **618,** 569–585 (2005).

43. Murray, N., Ménard, B. & Thompson, T. A. Radiation pressure from massive star clusters as a launching mechanism for super-galactic winds. *Astrophys. J.* **735,** 66 (2011).

44. Oppenheimer, B. D. & Davé, R. Cosmological simulations of intergalactic medium enrichment from galactic outflows. *Mon. Not. R. Astron. Soc.* **373,** 1265–1292 (2006).

45. Haas, M. R. *et al.* Physical properties of simulated galaxy populations at z = 2—I. Effect of metal-line cooling and feedback from star formation and AGN. *Mon. Not. R. Astron. Soc.* **435,** 2931–2954 (2013).

46. Kim, J.-h. *et al.* The AGORA high-resolution galaxy simulations comparison project. *Astrophys. J.* (Suppl.) **210,** 14 (2014).

47. Hopkins, P. F., Quataert, E. & Murray, N. Stellar feedback in galaxies and the origin of galaxy-scale winds. *Mon. Not. R. Astron. Soc.* **421,** 3522–3537 (2012).
This paper presents numerical simulations of feedback based on models of the underlying physical processes.

48. Booth, C. M. & Schaye, J. The interaction between feedback from active galactic nuclei and supernovae. *Sci. Rep.* **3,** 1738 (2013).

49. Hopkins, P. F. *et al.* Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation. *Mon. Not. R. Astron. Soc.* **445,** 581–603 (2014).

50. Martin, C. L. Mapping large-scale gaseous outflows in ultraluminous galaxies with Keck II ESI spectra: variations in outflow velocity with galactic mass. *Astrophys. J.* **621,** 227–245 (2005).
A demonstration of the scaling of outflow properties with galaxy properties in local starbursts.

51. Martin, C. L., Kobulnicky, H. A. & Heckman, T. M. The metal content of dwarf starburst winds: results from Chandra observations of NGC 1569. *Astrophys. J.* **574,** 663–692 (2002).

52. Strickland, D. K., Heckman, T. M., Colbert, E. J. M., Hoopes, C. G. & Weaver, K. A. A high spatial resolution X-ray and Hα study of hot gas in the halos of star-forming disk galaxies. II. Quantifying supernova feedback. *Astrophys. J.* **606,** 829–852 (2004).

53. Lehnert, M. D. & Heckman, T. M. Ionized gas in the halos of edge-on starburst galaxies: evidence for supernova-driven superwinds. *Astrophys. J.* **462,** 651 (1996).

54. Heckman, T. M., Lehnert, M. D., Strickland, D. K. & Armus, L. Absorption-line probes of gas and dust in galactic superwinds. *Astrophys. J.,* (Suppl.)*,* **129,** 493–516 (2000).

55. Bolatto, A. D. *et al.* Suppression of star formation in the galaxy NGC 253 by a starburst-driven molecular wind. *Nature* **499,** 450–453 (2013).

56. Heckman, T. M. *et al.* Extreme feedback and the epoch of reionization: clues in the local Universe. *Astrophys. J.* **730,** 5 (2011).
This paper reports on feedback in extreme galaxies in the local Universe which may be analogous to galaxies at much higher redshifts.

57. George, R. D. *et al.* Herschel reveals a molecular outflow in a z = 2.3 ULIRG. *Mon. Not. R. Astron. Soc.* **442,** 1877–1883 (2014).

58. Steidel, C. C. *et al.* Strong nebular line ratios in the spectra of z ~ 2–3 star forming galaxies: first results from KBSS-MOSFIRE. *Astrophys. J.* **795,** 165 (2014).

59. Sanders, R. L. *et al.* The MOSDEF Survey: mass, metallicity, and star-formation rate at z ~ 2.3. *Astrophys. J.* **799,** 138 (2015).

60. Wisnioski, E. *et al.* The KMOS³D Survey: design, first results, and the evolution of galaxy kinematics from 0.7 ≤ z ≤ 2.7. *Astrophys. J.* **799,** 209 (2015).

61. Pettini, M. *et al.* New observations of the interstellar medium in the Lyman break galaxy MS 1512-cB58. *Astrophys. J.* **569,** 742–757 (2002).

62. Quider, A. M., Pettini, M., Shapley, A. E. & Steidel, C. C. The ultraviolet spectrum of the gravitationally lensed galaxy 'the Cosmic Horseshoe': a close-up of a star-forming galaxy at z ≈ 2. *Mon. Not. R. Astron. Soc.* **398,** 1263–1278 (2009).

63. Dessauges-Zavadsky, M. *et al.* Rest-frame ultraviolet spectrum of the gravitationally lensed galaxy ''the 8 o'clock arc'': stellar and interstellar medium properties. *Astron. Astrophys.* **510,** A26 (2010).

64. Quider, A. M., Shapley, A. E., Pettini, M., Steidel, C. C. & Stark, D. P. A study of interstellar gas and stars in the gravitationally lensed galaxy 'the Cosmic Eye' from rest-frame ultraviolet spectroscopy. *Mon. Not. R. Astron. Soc.* **402,** 1467–1479 (2010).

65. Reddy, N. A. & Steidel, C. C. A steep faint-end slope of the UV luminosity function at z ≈ 2–3: implications for the global stellar mass density and star formation in low-mass halos. *Astrophys. J.* **692,** 778–803 (2009).
This paper shows that faint galaxies at z ≈ 2–3 make a notable contribution to the global star-formation density.

66. Oesch, P. A. *et al.* The evolution of the ultraviolet luminosity function from z ≈ 0.75 to z ≈ 2.5 using HST ERS WFC3/UVIS observations. *Astrophys. J.* **725,** L150–L155 (2010).

67. McLure, R. J. *et al.* A new multifield determination of the galaxy luminosity function at z = 7–9 incorporating the 2012 Hubble Ultra-Deep Field imaging. *Mon. Not. R. Astron. Soc.* **432,** 2696–2716 (2013).

68. Bouwens, R. J. *et al.* UV luminosity functions at redshifts z = 4 to z ≈ 10: 11000 galaxies from HST legacy fields. *Astrophys. J.* **803,** 34 (2015).
This paper shows that the faint-end slope of the luminosity function remains steep to the highest redshifts measured, indicating that faint galaxies are abundant.

69. Alavi, A. *et al.* Ultra-faint ultraviolet galaxies at z ~ 2 behind the lensing cluster A1689: the luminosity function, dust extinction, and star formation rate density. *Astrophys. J.* **780,** 143 (2014).

70. Atek, H. *et al.* New constraints on the faint end of the UV luminosity function at $z \sim 7$–8 using the gravitational lensing of the Hubble frontier fields cluster A2744. *Astrophys. J.* **800,** 18 (2015).
71. Fan, X., Carilli, C. L. & Keating, B. Observational constraints on cosmic reionization. *Annu. Rev. Astron. Astrophys.* **44,** 415–462 (2006).
72. Hinshaw, G. *et al.* Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results. *Astrophys. J.* (Suppl.) **208,** 19 (2013).
73. Planck Collaboration et al. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.* **571,** A16 (2014).
74. Wyithe, J. S. B. & Cen, R. The extended star formation history of the first generation of stars and the reionization of cosmic hydrogen. *Astrophys. J.* **659,** 890–907 (2007).
75. Kuhlen, M. & Faucher-Giguère, C.-A. Concordance models of reionization: implications for faint galaxies and escape fraction evolution. *Mon. Not. R. Astron. Soc.* **423,** 862–876 (2012).
76. Robertson, B. E. *et al.* New constraints on cosmic reionization from the 2012 Hubble ultra deep field campaign. *Astrophys. J.* **768,** 71 (2013).
    This paper discusses the need for faint galaxies in accomplishing reionization.
77. Nestor, D. B., Shapley, A. E., Steidel, C. C. & Siana, B. Narrowband imaging of escaping Lyman-continuum emission in the SSA22 field. *Astrophys. J.* **736,** 18 (2011).
78. Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C. & Siana, B. A refined estimate of the ionizing emissivity from galaxies at $z \cong 3$: spectroscopic follow-up in the SSA22a field. *Astrophys. J.* **765,** 47 (2013).
79. Mostardi, R. E. *et al.* Narrowband Lyman-continuum imaging of galaxies at $z \sim 2.85$. *Astrophys. J.* **779,** 65 (2013).
80. Erb, D. K. *et al.* Physical conditions in a young, unreddened, low-metallicity galaxy at high redshift. *Astrophys. J.* **719,** 1168–1190 (2010).
    This paper describes the spectral properties and physical conditions of a low-mass, low-metallicity galaxy at high redshift.
81. Stark, D. P. *et al.* Ultraviolet emission lines in young low-mass galaxies at $z \approx 2$: physical properties and implications for studies at $z > 7$. *Mon. Not. R. Astron. Soc.* **445,** 3200–3220 (2014).
82. Borthakur, S., Heckman, T. M., Leitherer, C. & Overzier, R. A. A local clue to the reionization of the universe. *Science* **346,** 216–219 (2014).
83. Hu, E. M. & McMahon, R. G. Detection of Lyman-α-emitting galaxies at redshift 4.55. *Nature* **382,** 231–233 (1996).
84. Gawiser, E. *et al.* The physical nature of Lyα-emitting galaxies at $z = 3.1$. *Astrophys. J.* **642,** L13–L16 (2006).
85. Guaita, L. *et al.* Lyα-emitting galaxies at $z = 2.1$: stellar masses, dust, and star formation histories from spectral energy distribution fitting. *Astrophys. J.* **733,** 114 (2011).
86. McLinden, E. M. *et al.* Galactic winds and stellar populations in Lyman α emitting galaxies at $z \approx 3.1$. *Mon. Not. R. Astron. Soc.* **439,** 446–473 (2014).
87. Vargas, C. J. *et al.* To stack or not to stack: spectral energy distribution properties of Lyα-emitting galaxies at $z = 2.1$. *Astrophys. J.* **783,** 26 (2014).
88. Chonis, T. S. *et al.* The spectrally resolved Lyα emission of three Lyα-selected field galaxies at $z \approx 2.4$ from the HETDEX pilot survey. *Astrophys. J.* **775,** 99 (2013).
89. Hashimoto, T. *et al.* Gas motion study of Lyα emitters at $z \approx 2$ using FUV and optical spectral lines. *Astrophys. J.* **765,** 70 (2013).
90. Guaita, L. *et al.* Magellan/MMIRS near-infrared multi-object spectroscopy of nebular emission from star-forming galaxies at $2 < z < 3$. *Astron. Astrophys.* **551,** A93 (2013).
91. Rhoads, J. E. *et al.* The dynamical masses, densities, and star formation scaling relations of Lyα galaxies. *Astrophys. J.* **780,** 20 (2014).
92. Song, M. *et al.* The HETDEX pilot survey. V. The physical origin of Lyα emitters probed by near-infrared spectroscopy. *Astrophys. J.* **791,** 3 (2014).
93. Shibuya, T. *et al.* What is thE PHYSICAL ORIGIN OF Strong Lyα emission? II. Gas kinematics and distribution of Lyα emitters. *Astrophys. J.* **788,** 74 (2014).
94. Erb, D. K. *et al.* The Lyα properties of faint galaxies at $z \sim 2$–3 with systemic redshifts and velocity dispersions from Keck-MOSFIRE. *Astrophys. J.* **795,** 33 (2014).
95. Verhamme, A., Schaerer, D. & Maselli, A. 3D Lyα radiation transfer. I. Understanding Lyα line profile morphologies. *Astron. Astrophys.* **460,** 397–413 (2006).
96. Verhamme, A., Schaerer, D., Atek, H. & Tapken, C. 3D Lyα radiation transfer. III. Constraints on gas and stellar properties of $z \approx 3$ Lyman break galaxies (LBG) and implications for high-z LBGs and Lyα emitters. *Astron. Astrophys.* **491,** 89–111 (2008).
97. Laursen, P., Sommer-Larsen, J. & Andersen, A. C. Lyα radiative transfer with dust: escape fractions from simulated high-redshift galaxies. *Astrophys. J.* **704,** 1640–1656 (2009).
98. Verhamme, A. *et al.* Lyman-α emission properties of simulated galaxies: interstellar medium structure and inclination effects. *Astron. Astrophys.* **546,** A111 (2012).
99. Kulas, K. R. *et al.* The kinematics of multiple-peaked Lyα emission in star-forming galaxies at $z \sim 2$–3. *Astrophys. J.* **745,** 33 (2012).
100. Law, D. R. *et al.* A HST/WFC3-IR morphological survey of galaxies at $z = 1.5$–3.6. II. The relation between morphology and gas-phase kinematics. *Astrophys. J.* **759,** 29 (2012).

ARTICLE

Glypican-1 identifies cancer exosomes and detects early pancreatic cancer

Sonia A. Melo[1]†*, Linda B. Luecke[1]*, Christoph Kahlert[1]*, Agustin F. Fernandez[2], Seth T. Gammon[3], Judith Kaye[1], Valerie S. LeBleu[1], Elizabeth A. Mittendorf[4], Juergen Weitz[5], Nuh Rahbari[5], Christoph Reissfelder[5], Christian Pilarsky[5], Mario F. Fraga[2,6], David Piwnica-Worms[3] & Raghu Kalluri[1]

Exosomes are lipid-bilayer-enclosed extracellular vesicles that contain proteins and nucleic acids. They are secreted by all cells and circulate in the blood. Specific detection and isolation of cancer-cell-derived exosomes in the circulation is currently lacking. Using mass spectrometry analyses, we identify a cell surface proteoglycan, glypican-1 (GPC1), specifically enriched on cancer-cell-derived exosomes. GPC1+ circulating exosomes (crExos) were monitored and isolated using flow cytometry from the serum of patients and mice with cancer. GPC1+ crExos were detected in the serum of patients with pancreatic cancer with absolute specificity and sensitivity, distinguishing healthy subjects and patients with a benign pancreatic disease from patients with early- and late-stage pancreatic cancer. Levels of GPC1+ crExos correlate with tumour burden and the survival of pre- and post-surgical patients. GPC1+ crExos from patients and from mice with spontaneous pancreatic tumours carry specific *KRAS* mutations, and reliably detect pancreatic intraepithelial lesions in mice despite negative signals by magnetic resonance imaging. GPC1+ crExos may serve as a potential non-invasive diagnostic and screening tool to detect early stages of pancreatic cancer to facilitate possible curative surgical therapy.

Exosomes are secreted membrane enclosed vesicles (extracellular vesicles) of 50–150 nm diameter[1]. Formed during the inward budding of late endosomes, they develop into intracellular multivesicular endosomes and contain nucleic acids and proteins[2–6]. Exosomes are released into the extracellular space and enter the circulation[7–10]. The biogenesis of exosomes is not clear, therefore the term extracellular vesicles is often used[11]. Exosomes-enriched proteins include members of the tetraspanin family (CD9, CD63 and CD81), members of the endosomal sorting complexes required for transport (TSG101 and Alix) and heat-shock proteins (Hsp60, Hsp70 and Hsp90)[12]. Specific markers that distinguish cancer exosomes from normal exosomes are unknown. Identification and isolation of cancer specific exosomes in body fluids could enable the identification of DNA, RNA and proteins without contamination from non-cancer exosomes, and aid in the treatment and management of cancer.

GPC1 is a specific marker of cancer exosomes

Extracellular vesicles from cancer cells (MDA-MB-231), fibroblasts (HDF and NIH/3T3) and non-tumorigenic cells (MCF10A and E10) were isolated by ultracentrifugation[13,14], and called exosomes based on the following observations. NanoSight nanoparticle tracking analysis and transmission electron microscopy (TEM) showed extracellular vesicles of 105 ± 5 nm (mean $\pm$ s.d.) and 112 ± 4 nm in diameter, respectively[15] (Extended Data Fig. 1a, b). Immunogold TEM (IG-TEM) showed CD9 (Extended Data Fig. 1c) and flotillin1 and CD81 by immunoblot[15] (Extended Data Fig. 1d and Extended Data Table 1a). Proteins were evaluated by ultra-performance liquid chromatography–mass spectrometry (UPLC–MS)[16] (Extended Data

Table 1a). Proteins from HDF, NIH/3T3, E10, MCF10A and MDA-MB-231 exosomes, included the exosomes markers TSG101, CD9 and CD63 (total number of proteins: HDF: 261, NIH/3T3: 171, E10: 232, MCF10A: 214 and MDA-MB-231: 242; Supplementary Table 1). Bioinformatic analyses revealed 48 proteins (25 cytoplasmic, 7 nuclear, 5 transmembrane, 1 membrane-anchored and 7 secreted) exclusively present in cancer exosomes (Fig. 1a, Extended Data Table 1a and Supplementary Table 1). Glypican-1 (GPC1) is a membrane-anchored protein that is overexpressed in breast and pancreatic cancer[17–19]. GPC1 was increased in breast and pancreatic cancer cell lines compared to non-tumorigenic cells (Extended Data Fig. 1e, f and Supplementary Fig. 1). Using immunoblot and IG-TEM, GPC1 was detected exclusively in cancer exosomes (Fig. 1, Extended Data Fig. 1g and Supplementary Fig. 1; HMLE cells).

We performed FACS analysis of exosomes to detect GPC1 protein (Fig. 1c). IG-TEM identified cancer exosomes with GPC1, while non-cancer exosomes did not exhibit GPC1 (Fig. 1d). Cancer exosomes from sucrose gradients or ultracentrifugation showed GPC1 (Fig. 1c, e, f, Extended Data Fig. 1h and Supplementary Fig. 1).

We implanted MDA-MB-231 cancer cells in the mammary fat pads of nude mice. The mice were bled before cancer cell inoculation, and again when tumours reached an average volume of 300, 550, 1,000 and 1,350 mm^3, and crExos were assessed for the presence of GPC1 (Extended Data Fig. 2a). The percentage of GPC1+ crExos increased proportionally with tumour size and correlated with tumour burden (Extended Data Fig. 2b, c). We stably expressed green fluorescent protein (GFP)-tagged CD63 in MDA-MB-231 cells. CD63 is an established exosomal marker[14], and MDA-MB-231-CD63–GFP-derived

[1]Department of Cancer Biology, Metastasis Research Center, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. [2]Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), HUCA, Universidad de Oviedo, 33006 Oviedo, Spain. [3]Department of Cancer Systems Imaging, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77054, USA. [4]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [5]Department of Gastrointestinal, Thoracic and Vascular Surgery, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Fetscherstr. 74, 01307 Dresden, Germany. [6]Department of Immunology and Oncology, National Center for Biotechnology, CNB-CSIC, Cantoblanco, 28049 Madrid, Spain. †Present address: Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal, Institute of Pathology and Molecular Immunology of the University of Porto (IPATIMUP), 4200 Porto, Portugal.
*These authors contributed equally to this work.
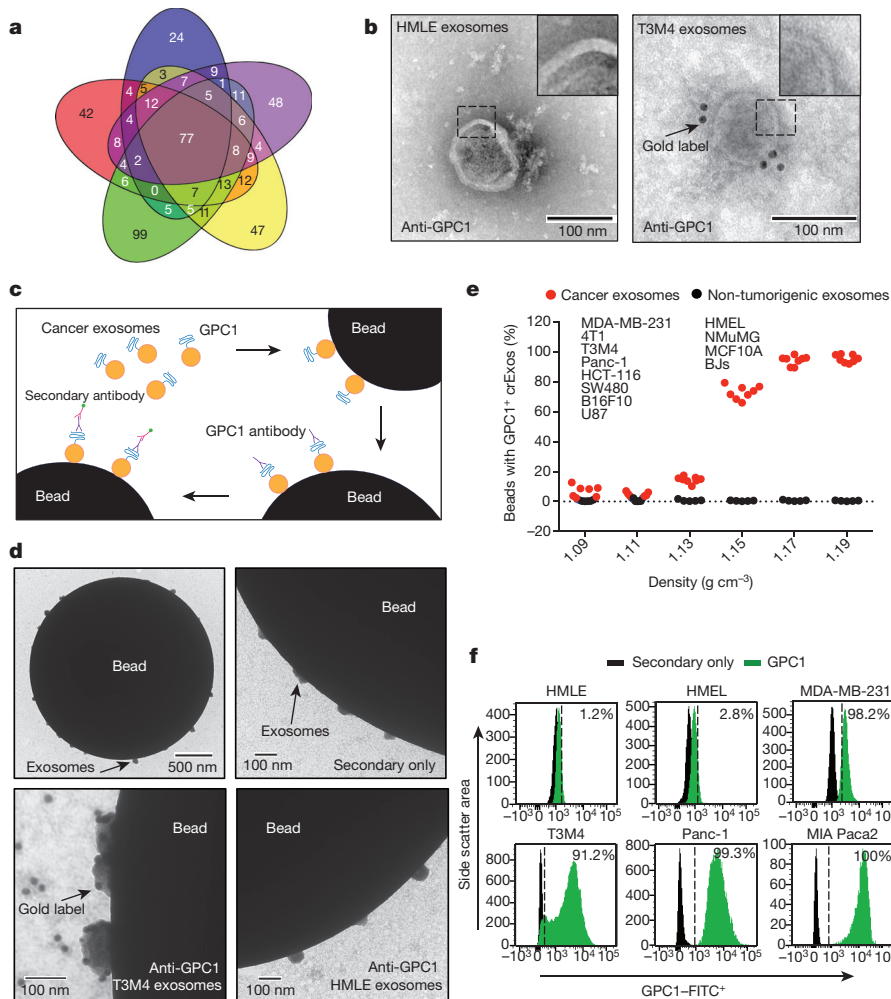
Figure 1 | **GPC1 is present on cancer exosomes.** **a**, Venn diagram of proteins from NIH/3T3 (blue), MCF10A (red), HDF (green), E10 (yellow) and MDA-MB-231 (purple) exosomes. In total, 48 proteins were exclusively detected in MDA-MB-231 exosomes ($n = 3$ protein samples, technical replicates). **b**, TEM (left) and IG-TEM (right) of GPC1. Top right, digitally zoomed inset ($n = 2$ experiments). **c**, Diagram of flow cytometry (FACS). **d**, TEM of bead-bound exosomes and immunogold labelling of GPC1 ($n = 2$ biological replicates). **e**, Percentage of GPC1+ exosome beads from cancer (red) and non-cancer cells (black). **f**, FACS of the percentage of GPC1+ exosome beads ($n = 2$ biological replicates). Negative control: secondary antibody only.

exosomes were uniformly positive for GFP (Extended Data Fig. 2d). crExos were also collected from mice with orthotopic MDA-MB-231-CD63–GFP tumours, and a subpopulation of the crExos were GFP+ (Extended Data Fig. 2e). GPC1 expression was exclusively detected in the GFP+ crExo fraction but not in GFP− crExos (Extended Data Fig. 2f).

GPC1+ circulating exosomes are a cancer biomarker

Next, we isolated crExos from patients with breast cancer ($n = 32$), pancreatic ductal adenocarcinoma (PDAC, $n = 190$) and healthy donors ($n = 100$) (patient data in Extended Data Table 2a). TEM analysis of crExos isolated from the serum revealed a lipid bilayer and CD9 positivity (Extended Data Fig. 3a, b). crExos from sucrose gradient purification also showed expression of the exosomes marker flotillin1 (Extended Data Fig. 3c, Supplementary Fig. 1)[14,15]. The relative concentration of crExos was significantly higher in the sera of cancer patients compared to healthy donors (Extended Data Fig. 3d), and the average size of PDAC crExos was significantly smaller than all other crExos (Extended Data Fig. 3e). Analyses of sera from healthy donors revealed baseline positivity for GPC1 in crExos, ranging from 0.3 to 4.7% (average of 2.3%; Fig. 2a). We observed that 75% of patients with breast cancer (24 out of 32) demonstrated GPC1+ crExos levels higher than healthy donors ($P < 0.0001$; Fig. 2a). Any specific correlation between the level of GPC1+ crExos and breast cancer subtypes was not appreciated in this patient cohort (luminal A, luminal B or triple-negative subtypes; Extended Data Fig. 3f). By contrast, all 190 PDAC patients revealed higher levels of GPC1+ crExos than in healthy donors ($P < 0.0001$; Fig. 2a and Extended Data Fig. 8a, b). These results indicated a strong correlation between GPC1+ crExos and cancer, particularly for PDAC.

GPC1+ crExos contain oncogenic KRAS[G12D]

Exosomes contain DNA and RNA[20]. KRAS is a frequently mutated gene in pancreatic cancer and mutant transcripts have been found in circulation[3,21,22]. Primary PDAC tumour samples of 47 PDAC patients were sequenced to assess KRAS status. Of these, 16 contained wild-type KRAS, 14 contained KRAS[G12D] (which encodes the KRAS glycine-to-aspartic-acid substitution mutant), 11 KRAS[G12V], 5 KRAS[G12R] and 1 KRAS[G12V/C] mutation (Fig. 2b). Sufficient amount of corresponding serum was available from 10 patients with KRAS[G12D] and 5 KRAS[G12V] mutations. IG-TEM in GPC1+ and GPC1− crExos from the same patient confirmed GPC1 presence (Fig. 2c). All 15 GPC1+ crExos with tumour validated KRAS mutation revealed identical mutation by quantitative PCR of exosomal messenger RNA (Fig. 2d). Wild-type KRAS mRNA was found both in GPC1+ and GPC1− crExos, while mutant KRAS transcript was only detected in the GPC1+ crExos (Fig. 2d).

GPC1+ crExos detect early pancreatic cancer

Analysis of the discovery cohort revealed that the levels of GPC1+ crExos distinguish patients with histologically validated pancreatic cancer precursor lesions (PCPL, $n = 5$; Extended Data Table 2a) from healthy donors and patients with benign pancreatic disease (BPD, $n = 26$; Extended Data Table 2a, b and Fig. 2e). Specifically, the levels of GPC1+ crExos in the PCPL group (intraductal papillary mucinous neoplasm (IPMN); $n = 5$) were consistently higher than the levels of GPC1+ crExos in the healthy donor group, as well as in the BPD group (which includes 18 patients with pancreatitis and 8 with cystic adenomas; Fig. 2e). All patients in PCPL group presented with specific clinical symptoms and exhibited a macroscopic mass using MRI or computed
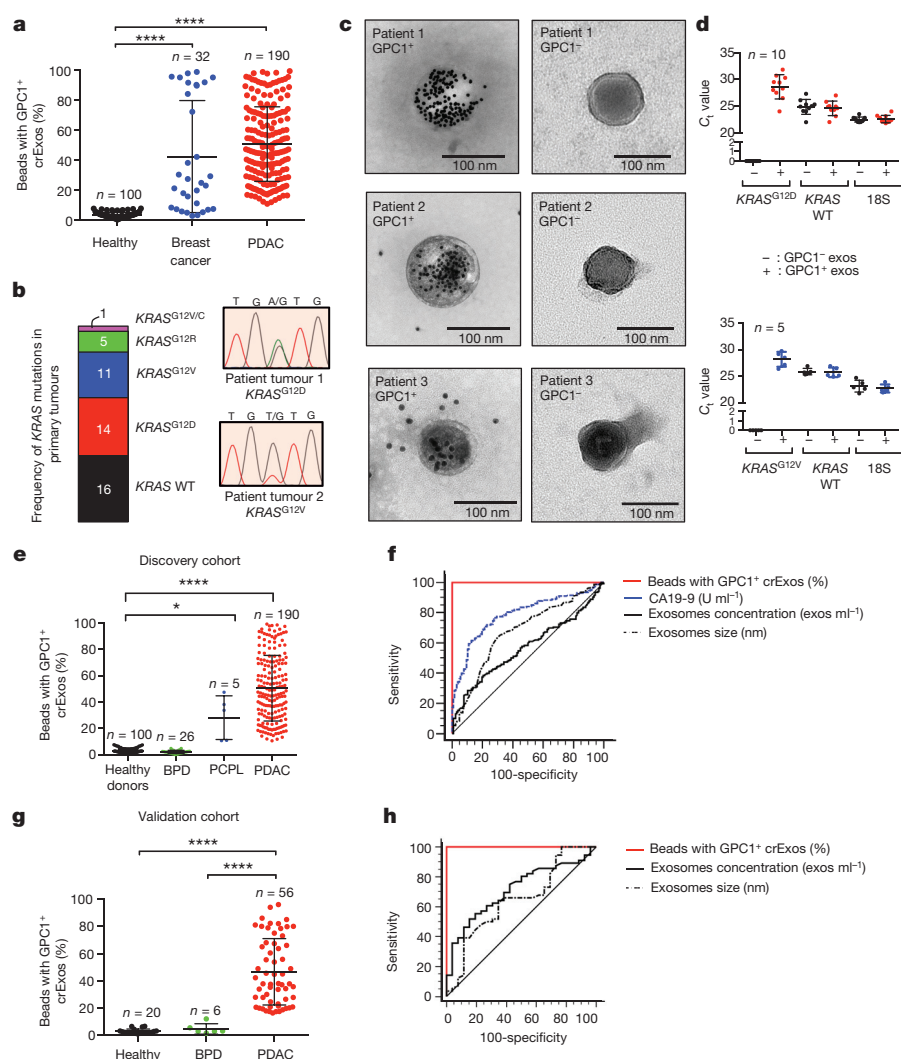
Figure 2 | GPC1$^+$ crExos are a non-invasive biomarker for pancreatic cancer. a, Percentage of GPC1$^+$crExo beads in healthy donors, patients with breast cancer and patients with PDAC (analysis of variance (ANOVA), post-hoc Tamhane T$_2$, ****$P < 0.0001$). **b**, Frequency of *KRAS* mutation in 47 tumours and representative DNA sequencing chromatograms. WT, wild type. **c**, IG-TEM of GPC1 of crExos from 3 PDAC patients following FACS isolation of GPC1$^+$ (left) and GPC1$^-$ (right) crExos ($n = 3$, 3 technical replicates). **d**, C_t value for *KRAS*$^{G12D/G12V}$, *KRAS* wild-type mRNA and 18S ribosomal RNA expression in GPC1$^+$ (red and blue) and GPC1$^-$ (black) crExos (after FACS isolation; $n = 2$, 2 biological replicates and 3 technical replicates each). **e**, Percentage of GPC1$^+$ crExo beads in the discovery cohort (ANOVA, post-hoc Tamhane T$_2$, **$P < 0.01$, ****$P < 0.0001$; 3 technical replicates). **f**, ROC curve of discovery cohort. **g**, Percentage of GPC1$^+$ crExo beads in the validation cohort (ANOVA, post-hoc Tukey-Kramer test, ****$P < 0.0001$; 3 technical replicates). **h**, ROC curve of validation cohort. Data are mean ± s.d.

tomography. The BPD group exhibited similar GPC1$^+$ crExos levels (average 2.1% GPC1$^+$ crExos) to healthy donors (Fig. 2e).

We compared the specificity and sensitivity of GPC1$^+$ crExos to levels of carbohydrate antigen 19-9 (CA19-9; also known as sialyl Lewis^A), the clinical standard tumour biomarker for patients with PDAC[23]. CA19-9 levels were increased in the serum of patients with PDAC when compared to healthy donors, but CA19-9 levels were also significantly increased in the serum of patients with BPD ($P < 0.0001$; Extended Data Fig. 3g). Notably, CA19-9 levels failed to distinguish patients with PCPL from healthy donors (Extended Data Fig. 3g). When comparing patients with stage I–IV pancreatic cancer to healthy donors and patients with BPD, the receiver operating characteristic (ROC) curves show that GPC1$^+$ crExos revealed a near perfect classifier with an AUC of 1.0 exhibiting a sensitivity and specificity of 100%, and with a positive and negative predictive value of 100% (Fig. 2f and Extended Data Fig. 4a–f). By contrast, CA19-9 was inferior in distinguishing patients with PDAC from healthy donors ($P < 0.001$; Fig. 2f and Extended Data Fig. 4a–f). Of note, neither the concentration of exosomes nor their size was a valid parameter to distinguish PDAC patients from controls (Fig. 2f and Extended Data Fig. 4a–f). GPC1$^+$ crExos showed a sensitivity and specificity of 100% in each stage of pancreatic cancer (carcinoma *in situ*, stage I as well as stages II–IV), supporting its utility as a biomarker for all stages of pancreatic cancer and its potential for early detection of the disease.

An independent patient cohort, composed of 6 patients with histologically validated BPD (chronic pancreatitis), 56 patients with PDAC and 20 healthy donors, was used to validate our findings

(Extended Data Table 2a). GPC1$^+$ crExos distinguished patients with PDAC from healthy donors and BPD patients (Fig. 2g). The BPD group exhibited similar GPC1$^+$ crExos levels to healthy donors (Fig. 2g). In complete agreement with the discovery cohort, ROC curves indicated that GPC1$^+$ crExos (from patients with PDAC or BPD and healthy donors) revealed a near perfect classifier with an AUC of 1.0, and sensitivity, specificity, positive and negative predictive values of 100% (Fig. 2h and Extended Data Fig. 4g).

GPC1$^+$ crExos inform pancreatic cancer burden

We next sought to evaluate whether GPC1$^+$ crExo levels could inform on metastatic disease burden of PDAC patients (Extended Data Table 2a). GPC1$^+$ crExos of PDAC patients with distant metastatic disease showed significantly higher levels of bead-bound GPC1$^+$ crExos (average 58.5%) than patients with metastatic disease restricted to lymph nodes (average 50.5%) or no metastases (average 39.9%; Extended Data Fig. 5a). Furthermore, we evaluated GPC1$^+$ crExos in serum of PDAC patients at pre- and post-surgery stages (post-operative day 7; PDAC $n = 29$, PCPL $n = 4$ and BPD $n = 4$; Fig. 3a). In total, 28 out of 29 PDAC patients and all PCPL patients with longitudinal blood collections showed a significant decrease in GPC1$^+$ crExo levels after surgical resection (PDAC: $P < 0.0001$; PCPL: $P < 0.001$; Fig. 3b). By contrast, CA19-9 levels decreased in only 19 out of 29 PDAC patients, and in none of the PCPL patients (PDAC: $P = 0.003$; PCPL: $P = 0.81$; Extended Data Fig. 5b). In 4 BPD patients, neither the levels of GPC1$^+$ crExos nor the levels of CA19-9 showed a difference (Fig. 3b and Extended Data Fig. 5b).
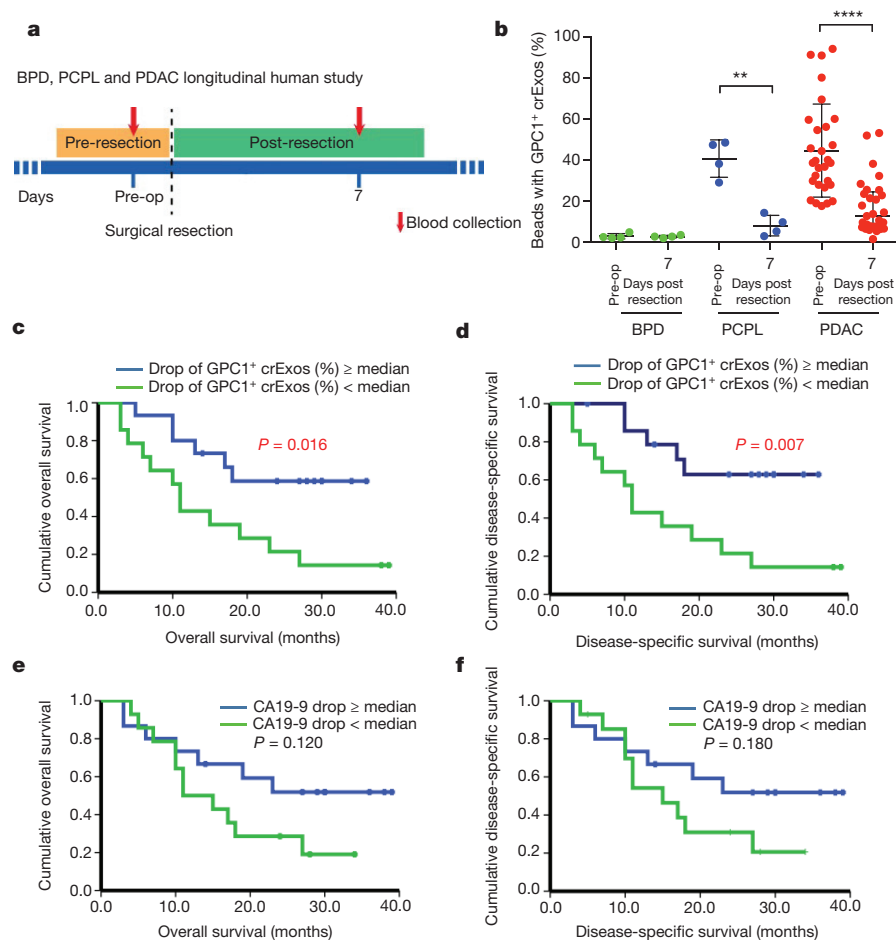
Figure 3 | Levels of circulating GPC1$^+$ exosomes inform pancreatic cancer resection outcome.
a, Longitudinal blood collection pre- and post-operatively (day 7). **b**, Percentage of GPC1$^+$crExo beads from patients with BPD ($n = 4$), PCPL ($n = 4$) or PDAC ($n = 29$) (paired two-tailed Student's t-test, **$P < 0.01$, ****$P < 0.0001$). Data are mean $\pm$ s.d. **c**, **d**, Kaplan–Meier curves (log-rank test) displaying overall (**c**) and disease-specific (**d**) survival of patients with a GPC1$^+$crExo drop $\geq$ median drop (blue), and with a GPC1$^+$ crExo drop $<$ median drop (green) after resection. **e**, **f**, Kaplan–Meier curves (log-rank test) displaying overall (**e**) and disease-specific (**f**) survival of patients with a CA19-9 drop $\geq$ median (blue), and a CA19-9 drop $<$ median (green) drop after resection.

To determine the prognostic relevance of GPC1$^+$ crExos in this longitudinal study (Fig. 3a), patients were dichotomized into two groups. Group 1 was defined by a decrease of GPC1$^+$ crExos greater or equal to the median decrease in GPC1$^+$ crExos, and group 2 was defined by a decrease of GPC1$^+$crExos that was less than the median decrease of GPC1$^+$crExos. Group 1 presented with improved overall (26.2 months) and disease-specific (27.7 months) survival when compared to group 2 (15.5 months for both overall and disease-specific; Fig. 3c, d). Although a decrease in CA19-9 levels is noted, this decrease did not significantly associate with overall and disease-specific survival (Fig. 3e, f and Extended Data Fig. 5b). Using a Cox regression model for a multivariate analysis confirmed the decrease in GPC1$^+$crExos, as an independent prognostic and predictive marker for disease-specific survival (Extended Data Fig. 5c, d).

Next, we evaluated whether an ELISA for circulating GPC1 could function with the same specificity and sensitivity as GPC1$^+$ crExos. Serum samples of the validation cohort (20 healthy donors, 6 BPD and 56 PDAC patients) were analysed for circulating GPC1 levels. While GPC1 levels were significantly higher in patients with PDAC than in patients with BPD and healthy donors, the sensitivity and specificity of the assay was lower when compared to GPC1$^+$ crExos. The GPC1 ELISA was similar to circulating CA19-9 assay. ROC curves indicated that circulating GPC1 protein shows an AUC of 0.781 a sensitivity of 82.14%, a specificity of 75%, and positive and negative predictor values of 4% and 100%, respectively (Extended Data Fig. 5e, f).

GPC1$^+$ crExos detect early PanIN lesions

We next evaluated the longitudinal appearance in the serum of GPC1$^+$crExos relative to pancreatic tumour burden. To this end, we used a genetically engineered mouse model (GEMM) for PDAC. The $Ptf1a^{cre/+}; LSL\text{-}Kras^{G12D/+}; Tgfbr2^{L/L}$ (PKT) mice[24] develop PDAC with

full penetrance that reliably recapitulates the clinical and histopathological features of the human disease[24,25]. The PKT mice consistently progress from pancreatic intraepithelial neoplasia (PanIN) at 4.5 weeks of age and die at 8 weeks of age owing to PDAC[24,25] (Extended Data Fig. 6a). In a longitudinal study, we bled PKT and littermate control mice repeatedly at 4, 5, 6, 7 and 8 weeks of age ($n = 7$ PKT mice and $n = 6$ control mice; Fig. 4a). Then 3 out of 7 PKT mice were euthanized by week 7, along with 4 out of 6 controls, while the remaining 3 PKT mice and 2 controls were euthanized at week 8. At 4 weeks of age, PKT mice showed on average an 8.4% elevation in GPC1$^+$ crExos, and this increased proportionally with time (and tumour burden) and severity of disease (histopathology), whereas control mice showed an average of 1.2% GPC1$^+$ crExos and this level remained constant with time (Fig. 4b). crExo sizes and concentration did not consistently correlate with disease over time (Extended Data Fig. 6b, c). MRI[26] was performed at the same time points when mice were bled to measure GPC1$^+$ crExos (Fig. 4c). When evaluated as a group, GPC1$^+$ crExo levels appeared before MRI detectable pancreatic masses (Fig. 4c, d and Extended Data Fig. 6d). GPC1$^+$ crExo size and concentration minimally correlated with pancreatic cancer (Extended Data Fig. 6b, c), whereas GPC1$^+$crExo levels correlated with tumour volume determined by MRI, and appeared to lead the growth of the tumour (Pearson correlation test, $r = 0.67$, $P = 0.0005$; Fig. 4c, d and Extended Data Fig. 6d). Notably, no increase in GPC1$^+$ crExo levels was noted in a mouse model of cerulein-induced acute pancreatitis, supporting the idea that the GPC1$^+$ crExo increase is pancreatic-cancer-specific (Extended Data Fig. 6e). The ROC curve of GPC1$^+$crExos showed an AUC of 1.0 in PKT compared to healthy littermate control mice at all ages evaluated (Fig. 4e and Extended Data Fig. 6f).

A cross-sectional study assayed tumour burden and GPC1$^+$ crExos in PKT mice, as early as 16 and 20 days of age, when they present with
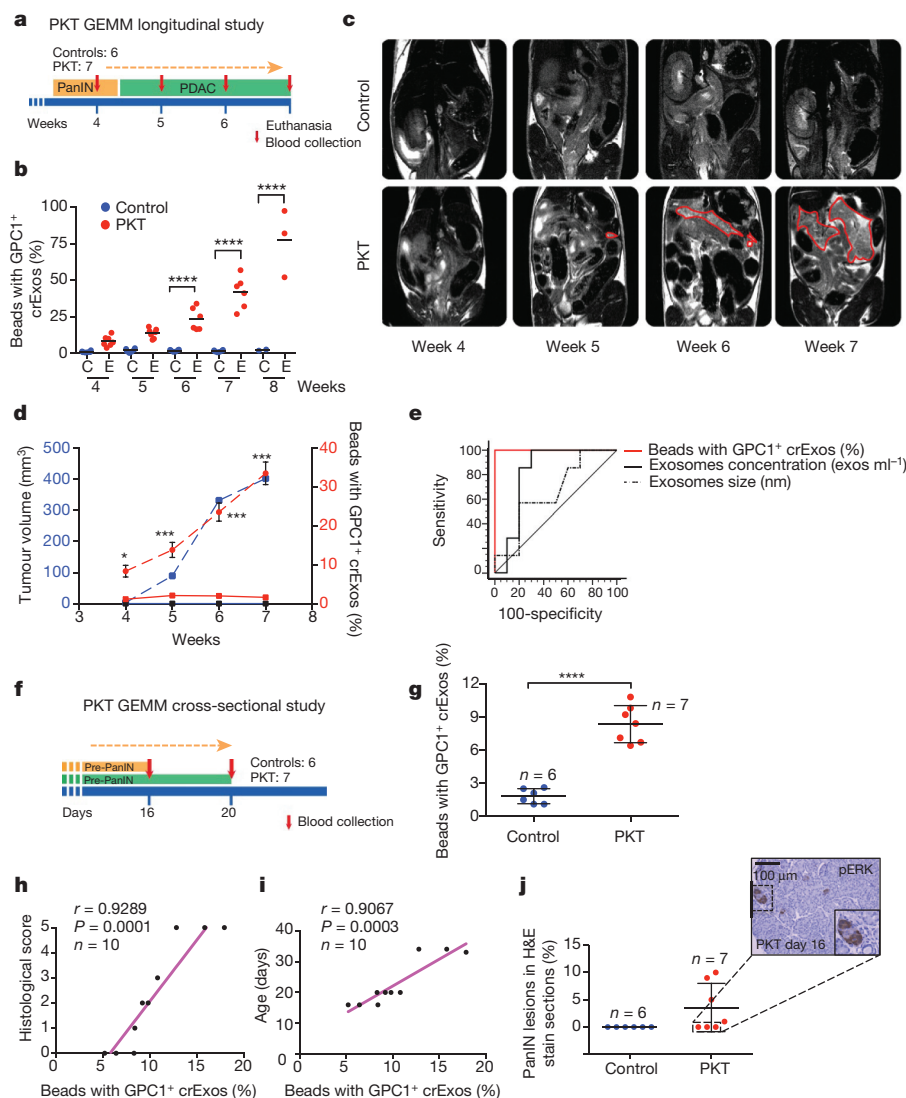
Figure 4 | GPC1$^+$ circulating exosomes predict pancreatic cancer in GEMMs. a, Longitudinal blood collection of control and PKT mice, at 4 ($n = 6$ and $n = 7$, respectively), 5 ($n = 6$ and $n = 7$), 6 ($n = 6$ and $n = 6$), 7 ($n = 6$ and $n = 6$) and 8 ($n = 2$ and $n = 3$) weeks of age. **b**, Percentage of GPC1$^+$ crExo beads from PKT (red) and control (blue) mice (ANOVA, post-hoc Tukey–Kramer test, ****$P < 0.0001$; 3 technical replicates). **c**, MRI with tumour encircled in red. **d**, Tumour volume and percentage of GPC1$^+$ crExos in PKT mice at indicated age (ANOVA, post-hoc Tamhane T$_2$, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ ****$P < 0.0001$; 3 technical replicates). **e**, ROC curve analysis of 4-week-old control mice ($n = 6$) and mice with acute pancreatitis ($n = 4$) versus 4-week-old PKT mice ($n = 7$). **f**, Cross-sectional study; blood collected from 16- or 20-day-old control ($n = 6$) and PKT ($n = 7$) mice. **g**, Percentage of GPC1$^+$ crExo beads from control and PKT (16–20-day-old) mice (paired two-tailed Student's t-test, ****$P < 0.0001$; 3 technical replicates). **h**, Graphical representation of correlation between histopathological score and GPC1$^+$ crExo levels. **i**, Graphical representation of correlation between age of PKT mice and GPC1$^+$ crExo levels. **j**, Relative percentage of PanIN lesions and representative haematoxylin and eosin (H&E) staining for phosphorylated ERK. Data are mean $\pm$ s.d.

pre-PanIN to early PanIN lesions (Fig. 4f, Extended Data Fig. 7a and Extended Data Table 3). GPC1$^+$crExos were detected in all PKT mice (PKT: 8.3% average, control: 1.8% average; Fig. 4g and Extended Data Table 1b). Histological analysis of PKT mice confirmed pre-PanIN lesions in 4 out of 7 PKT mice, and despite no observed histological lesions in 3 out of 7 PKT mice, GPC1$^+$ crExos predicted future pancreatic cancer emergence (Extended Data Table 1b). Moreover, we did not observe pancreas-associated masses by MRI in 16- and 20-day-old PKT mice. Both the histopathological score and age of PKT mice correlated with GPC1$^+$ crExo levels (Fig. 4h, i). In 4 out of 7 PKT mice with no observed histological lesions, downstream signals for Kras activation, such as phosphorylated ERK (pERK), were detected in the pancreas tissue (Fig. 4j and Extended Data Fig. 7a). We also observed exclusive detection of mutant $Kras^{G12D}$ mRNA in GPC1$^+$ crExos compared to GPC1$^-$ crExos (Extended Data Fig. 7b).

Discussion

Tumour exosomes are enriched in GPC1, and GPC1$^+$ crExos contain mutant $KRAS$ mRNA. We show that GPC1$^+$crExos are a reliable biomarker for detection of early pancreatic cancer. GPC1$^+$crExos are a prognostic marker superior to CA19-9. GPC1$^+$ crExos lead MRI as they can be detected in circulation before MRI-detectable lesions in GEMM of pancreatic cancer.

Routine screening for PDAC using MRI or computed tomography would be prohibitively expensive and associated with a high false-positive rate[27]. GPC1$^+$ crExos detect the possibility of pancreatic cancer in 16-day-old mice with unremarkable pancreatic histology and negative MRI. These results suggest the use of GPC1$^+$ crExos as a detection and monitoring tool for pancreatic cancer, with an emphasis in early detection.

Although $KRAS$ mutations are likely driver mutations for pancreatic cancer and are detected in early PanIN-1 lesions, it is estimated that 15–20 years may lapse before early PanIN lesions become metastatic PDAC[28–30]. Nonetheless, PDAC currently presents late with nonspecific clinical symptoms, therefore, as many as 80% of patients present with metastasis at diagnosis[31]. Patients with pancreatic cancer exhibit increased serum levels of CA19-9, carcinoembryonic antigen, CA-50, SPan-1, peanut agglutinin, DU-PAN-2, a-fetoprotein, tissue polypeptide antigen and pancreatic oncofetal antigen[32]. While these markers exhibit some use in tracking biopsy-diagnosed disease, they are also increased in patients with BPD. The lack of specific serum biomarkers and retroperitoneal position of the pancreas challenges the early detection of pancreatic cancer[33,34]. Pancreatico-duodenectomy can be curative if tumours are detected early[35]. Owing to the late diagnosis of pancreatic cancer, only around 15% of patients present with surgically resectable tumours[36]. Studies comparing stage of disease with outcome after surgery suggest that death rates would be reduced if the disease were diagnosed at earlier stages[37].

The isolation of cancer exosomes from patients remains a challenge owing to the lack of specific markers that can distinguish cancer from non-cancer exosomes. Genetic profiling on circulating DNA is cofounded by the fact that the isolated DNA has tumour and non-

tumour origins, thus making mutation detection challenging[38,39]. We previously demonstrated that the DNA in the circulation is mainly associated with exosomes[6]. Therefore, a marker for cancer exosomes will increase the sensitivity of detection for low frequency mutations in the circulation. As a proof-of-concept, GPC1[+]crExos identified *KRAS* mutations with 100% correlation with *KRAS* mutations in the tumour.

Our results provide evidence for GPC1 as a pan-specific marker of cancer exosomes. GPC1 is a proteoglycan that interacts with many proteins and has diverse functions[40]. Many cancer cells overexpress GPC1, with the most abundant increases observed in pancreatic cancer cells lines and tissue[17–19]. Studies have suggested a role for GPC1 as a positive regulator of cancer progression using orthotopic and GEMMs of PDAC[41,42]. GPC1 is an attractive candidate for detection and isolation of exosomes in the circulation of patients with cancer for genetic and proteomic analysis of specific alterations. Such opportunity offers the possibility for early detection of pancreatic cancer and help in designing potential curative surgical options.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Pan, B. T., Teng, K., Wu, C., Adam, M. & Johnstone, R. M. Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. *J. Cell Biol.* **101,** 942–948 (1985).
2. Trajkovic, K. *et al.* Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319,** 1244–1247 (2008).
3. Skog, J. *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature Cell Biol.* **10,** 1470–1476 (2008).
4. Al-Nedawi, K. *et al.* Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nature Cell Biol.* **10,** 619–624 (2008).
5. Demory Beckler, M. *et al.* Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. *Mol. Cell Proteomics* **12,** 343–355 (2013).
6. Kahlert, C. *et al.* Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J. Biol. Chem.* **289,** 3869–3875 (2014).
7. Ostrowski, M. *et al.* Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nature Cell Biol.* **12,** 19–30 (2010).
8. Théry, C., Ostrowski, M. & Segura, E. Membrane vesicles as conveyors of immune responses. *Nature Rev. Immunol.* **9,** 581–593 (2009).
9. Janowska-Wieczorek, A. *et al.* Microvesicles derived from activated platelets induce metastasis and angiogenesis in lung cancer. *Int. J. Cancer* **113,** 752–760 (2005).
10. Hergenreider, E. *et al.* Atheroprotective communication between endothelial cells and smooth muscle cells through miRNAs. *Nature Cell Biol.* **14,** 249–256 (2012).
11. Lotvall, J. *et al.* Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J. Extracell. Vesicles* **3,** 26913 (2014).
12. Taylor, D. D. & Gercel-Taylor, C. Exosomes/microvesicles: mediators of cancer-associated immunosuppressive microenvironments. *Semin. Immunopathol.* **33,** 441–454 (2011).
13. Luga, V. *et al.* Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* **151,** 1542–1556 (2012).
14. Théry, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* **Chapter 3,** Unit–3.22 (2006).
15. Théry, C., Zitvogel, L. & Amigorena, S. Exosomes: composition, biogenesis and function. *Nature Rev. Immunol.* **2,** 569–579 (2002).
16. Wilson, I. D. *et al.* High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* **4,** 591–598 (2005).
17. Matsuda, K. *et al.* Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer Res.* **61,** 5562–5569 (2001).
18. Kleeff, J. *et al.* The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *J. Clin. Invest.* **102,** 1662–1673 (1998).
19. Su, G. *et al.* Glypican-1 is frequently overexpressed in human gliomas and enhances FGF-2 signaling in glioma cells. *Am. J. Pathol.* **168,** 2014–2026 (2006).
20. Kahlert, C. & Kalluri, R. Exosomes in tumor microenvironment influence cancer progression and metastasis. *J. Mol. Med.* **91,** 431–437 (2013).
21. Morris, J. P., t., Wang, S. C. & Hebrok, M. KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. *Nature Rev. Cancer* **10,** 683–695 (2010).
22. Chen, W. W. *et al.* BEAMing and droplet digital PCR analysis of mutant IDH1 mRNA in glioma patient serum and cerebrospinal fluid extracellular vesicles. *Mol. Ther. Nucleic Acids* **2,** e109 (2013).
23. Del Villano, B. C. *et al.* Radioimmunometric assay for a monoclonal antibody-defined tumor marker, CA 19–9. *Clin. Chem.* **29,** 549–552 (1983).
24. Özdemir, B. C. *et al.* Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* **25,** 719–734 (2014).
25. Ijichi, H. *et al.* Aggressive pancreatic ductal adenocarcinoma in mice caused by pancreas-specific blockade of transforming growth factor-β signaling in cooperation with active Kras expression. *Genes Dev.* **20,** 3147–3160 (2006).
26. Lee, E. S. & Lee, J. M. Imaging diagnosis of pancreatic cancer: A state-of-the-art review. *World J. Gastroenterol.* **20,** 7864–7877 (2014).
27. Rickes, S., Unkrodt, K., Neye, H., Ocran, K. W. & Wermke, W. Differentiation of pancreatic tumours by conventional ultrasound, unenhanced and echo-enhanced power Doppler sonography. *Scand. J. Gastroenterol.* **37,** 1313–1320 (2002).
28. Murphy, S. J. *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterol.* **145,** 1098–1109 (2013).
29. Bardeesy, N. & DePinho, R. A. Pancreatic cancer biology and genetics. *Nature Rev. Cancer* **2,** 897–909 (2002).
30. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467,** 1114–1117 (2010).
31. Hidalgo, M. Pancreatic cancer. *N. Engl. J. Med.* **362,** 1605–1617 (2010).
32. Ballehaninna, U. K. & Chamberlain, R. S. Biomarkers for pancreatic cancer: promising new markers and options beyond CA 19–9. *Tumour Biol.* **34,** 3279–3292 (2013).
33. Jazieh, K. A., Foote, M. B. & Diaz, L. A. Jr. The clinical utility of biomarkers in the management of pancreatic adenocarcinoma. *Semin. Radiat. Oncol.* **24,** 67–76 (2014).
34. Locker, G. Y. *et al.* ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clinical Oncol.* **24,** 5313–5327 (2006).
35. Okano, K. & Suzuki, Y. Strategies for early detection of resectable pancreatic cancer. *World J. Gastroenterol.* **20,** 11230–11240 (2014).
36. Conlon, K. C., Klimstra, D. S. & Brennan, M. F. Long-term survival after curative resection for pancreatic ductal adenocarcinoma. Clinicopathologic analysis of 5-year survivors. *Ann. Surg.* **223,** 273–279 (1996).
37. Bilimoria, K. Y. *et al.* National failure to operate on early stage pancreatic cancer. *Ann. Surg.* **246,** 173–180 (2007).
38. Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497,** 108–112 (2013).
39. Yong, E. Cancer biomarkers: Written in blood. *Nature* **511,** 524–526 (2014).
40. David, G. *et al.* Molecular cloning of a phosphatidylinositol-anchored membrane heparan sulfate proteoglycan from human lung fibroblasts. *J. Cell Biol.* **111,** 3165–3176 (1990).
41. Whipple, C. A., Young, A. L. & Korc, M. A KrasG12D-driven genetic mouse model of pancreatic cancer requires glypican-1 for efficient proliferation and angiogenesis. *Oncogene* **31,** 2535–2544 (2012).
42. Aikawa, T. *et al.* Glypican-1 modulates the angiogenic and metastatic potential of human and mouse cancer cells. *J. Clin. Invest.* **118,** 89–99 (2008).

Author Contributions S.A.M. conceptually designed and carried out most of the experiments, generated the data and the figures, and wrote the manuscript; L.B.L. carried out most of the experiments; C.K. performed sequencing of primary tumours, performed most of the statistical analysis of the manuscript and supported manuscript writing; A.F.F. optimized, performed and analysed the UPLC–MS data; S.T.G. performed mouse MRI and analysed the data; J.K. performed the GEMM breeding, bleeding, euthanasia and material collection; V.S.L. performed the GEMM breeding, bleeding, euthanasia, material collection and supported manuscript editing; E.A.M. provided breast cancer patient samples and patient history; J.W., N.R., C.R. and C.P. collected and provided serum samples and patient history for analysis; M.F.F. optimized, performed and analyzed the UPLC–MS data; D.P.-W. performed the mouse MRI interpretation and provided support in data interpretation and analysis; R.K. conceived the idea, conceptually designed the study, supervised the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.K. (rkalluri@mdanderson.org).

METHODS

Patient samples and tissue collection. The study was conducted according to the Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK) criteria. The studies using human samples were designed as an explorative study. As there was no interventional approach in this study, a priori power calculation was not applicable. Instead, the number of patients included was assessed based on previous studies investigating the diagnostic relevance of circulating biologicals in pancreatic cancer[43].

Serum samples and tissue samples from patients with pancreatic cancer, serum samples only from patients with a benign pancreatic disease and from healthy donors, who had no evidence of acute or chronic or malignant disease and had no surgery within the past 12 months, were received from the department of General, Visceral and Transplantation Surgery from the University of Heidelberg and from the University Hospital of Dresden after approval by the local Institutional Review Board (IRB; Heidelberg: 323/2004, Dresden: EK357112012). The cases were obtained under an IRB-exempt protocol of the MD Anderson Cancer Center (IRB no. PA14-0154). Serum samples from patients with breast cancer were collected at the MD Anderson Cancer Center after approval of the Institutional Review Board (IRB no. LAB10-0690). A written consent for the serum sampling and tumour sampling was obtained pre-operatively from all patients and before serum collection from each healthy donor with disclosure of planned analyses regarding potential prognostic markers. The patients included in this study were all consecutive patients who underwent a surgical procedure at the University Hospital of Heidelberg, Germany, at the University Hospital of Dresden, Germany (pancreatic disease) or at the MD Anderson Cancer Center (breast cancer). All samples were randomly selected from larger cohorts and were analysed in a blinded fashion. Unblinding of clinical parameters and corresponding experimental data was performed only after finishing all experiments. Inclusion criteria of patients were a minimum of 18 years of age, histologically verified pancreatic cancer (pancreatic ductal adenocarcinoma), histologically verified benign pancreatic disease or breast cancer in a resection specimen, and a negative medical history for any other malignant disease. All blood samples were taken before treatment. Inclusion criteria for healthy control donors were a negative medical history for any malignant disease.

On the day of surgery, 10 ml serum separator tubes were used to collect blood samples before surgical incision. The blood samples were then centrifuged at 2,500g for 10 min to extract the serum, and the serum was stored at $-80\,^{\circ}$C until analysed. Likewise, blood samples were collected on day 7 after surgery for 29 patients with PDAC, 4 patients with chronic pancreatitis and 4 patients with an IPMN.

Patient characteristics and clinical specimens. The pancreatic discovery cohort from the University Hospital of Heidelberg included 190 patients with a PDAC, 18 patients with pancreatitis, 8 patients with a benign serous cystadenoma and 5 patients with IPMN. Patients were subjected to surgery between 2006 and 2012 at the Department of General, Visceral, and Transplantation Surgery, University of Heidelberg. Clinical information included age, gender, American Joint Committee on Cancer (AJCC) tumour stage, tumour size (pT), presence and number of lymph node metastases (pN), tumour grade (G), and treatment with (neo-)/adjuvant chemotherapy. The pancreatic cohort from the University Hospital of Dresden included 56 patients with PDAC, 6 patients with chronic pancreatitis, and 20 healthy donors. Patients were subjected to surgery between 2007 and 2013 at the Department of Gastrointestinal, Thoracic and Vascular Surgery, University of Dresden. Clinical information included age, gender, AJCC tumour stage, tumour size (pT), presence and number of lymph node metastases (pN), tumour grade (G), and treatment with (neo-)/adjuvant chemotherapy. The breast cancer cohort consisted of 32 women with breast cancer. All breast cancer patients were treated at the MD Anderson Cancer Center, Houston, Texas. Clinical information included age, gender, AJCC tumour stage, tumour size (pT), presence and number of lymph node metastases (pN), tumour grade, and treatment with (neo-)/adjuvant chemotherapy.

Animal studies. Female nude mice (*nu/nu*) (purchased from Jackson Laboratory) underwent breast pad injections with 0.5 million MDA-MB-231 cells or MDA-MB-231-CD63GFP cells in 20 μl of PBS injected per breast pad. Blood was collected retro-orbitally and exosomes were isolated before injection and at tumour volumes of 300, 550 1,000 and 1,350 mm^3. Mice were euthanized when the tumour size reached 1,500 mm^3 or when severe disease symptoms were present.

The disease progression and genotyping for the *Ptf1a*$^{cre/+}$; *LSL-Kras*$^{G12D/+}$; *Tgfbr2*$^{L/L}$ (PKT) and the *Pdx1*$^{cre/+}$; *LSL-Kras*$^{G12D/+}$; *p53*$^{R172H/+}$ (KPC) mice was previously described[32,33] (total of 13 females and 20 males mice). In the PKT longitudinal cohort, retro-orbital blood collections were performed at 4, 5, 6, 7 and 8 weeks of age. Mice were euthanized at 8 weeks of age or sooner if severe disease symptoms were noted. Histopathological analysis of mouse pancreas

specimen was performed following previously defined criteria[4]. Four C57BL/6 adult mice were subjected to repeated cerulean injection to induce acute pancreatitis (five hourly repeated intraperitoneal injections of 50 μg cerulein per kilogram of body weight) and euthanized 24 h after injection the last injection. Histological analyses of pancreas of mice was performed according to ref. 44, and a histological score was attributed according to the type of lesions detected: score 1: PanIN1a, score 2: PanIN1 a/b, score 3: PanIN2, score 4: PanIN3, score 5: ductal adenocarcinoma. All mice were housed under standard housing conditions at the MD Anderson Cancer Center (MDACC) animal facilities, and all animal procedures were reviewed and approved by the MDACC Institutional Animal Care and Use Committee.

Cell lines. The following human cells lines were used: HMLE (American Type Culture Collection (ATCC)), MCF10-A (human mammary epithelial cells, ATCC), BJ (ATCC), HDF (human dermal fibroblasts, ATCC), HMEL (human mammary epithelial cells, ATCC), MCF-7 (ATCC), MDA-MB-231 (triple-negative human metastatic breast carcinoma, ATTC), Panc-1 (ATTC), HCT-116 (ATCC), SW480 (ATCC), MIA Paca2 (ATCC) and T3M4 cells (Cell Bank, RIKEN BioResource Centre). The following murine cells lines were used: NIH/3T3 (mouse embryonic fibroblasts, ATCC), E10 (mouse lung epithelial cells, ATCC), NMuMG (ATTC), 4T1 (ATTC) and B16F10 cells (ATTC). All cell lines have been tested for mycoplasma contamination. HDF cells were cultured in DMEM supplemented with 20% (v/v) FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. HMLE cells and MCF10A cells were grown in DMEM/F12 supplemented with 5% (v/v) horse serum, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, 20 ng ml^{-1} EGF, 0.5 mg ml^{-1} hydrocortisone, 100 ng ml^{-1} cholera toxin and 10 μg ml^{-1} insulin. HMEL, MCF-7, MDA-MB-231, HCT-116, SW480, 4T1, NIH/3T3, E10, U87 and B16F10 cells were maintained in DMEM supplemented with 10% (v/v) FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. Panc-1, MIA Paca2 and T3M4 cells were cultured in RPMI-1640 supplemented with 10% (v/v) FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. NMuMG cells were grown in DMEM supplemented with 10% (v/v) FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin and 10 μg ml^{-1} insulin. All cell lines were kept in a humidifying atmosphere at 5% CO$_2$ at 37 °C. MDA-MB-231-CD63–GFP cells were engineered by transfection with a plasmid encoding a CD63–GFP fusion protein expressed under the control of a CMV promoter (p-CMV6-CD63-GFP from Origene, RG217238). Transfections were performed using Lipofectamine 2000 reagent (Invitrogen).

Exosomes isolation from cells. Exosomes were obtained from supernatant of cells as previously described[6] with some modifications[6]. In brief, cells were grown in T225 cm^2 flasks in FBS depleted of exosomes RPMI media until they reached a confluency of 80–90%. Next, the media was collected and centrifuged at 800g for 5 min, followed by a centrifugation step of 2,000g for 10 min to discard cellular debris. Then, the media was filtered using a 0.2-μm pore filter (syringe filter, 6786-1302, GE Healthcare). The collected media was then ultracentrifuged at 100,000g for 2 h at 4 °C. The exosomes pellet was washed with 35 ml PBS, followed by a second step of ultracentrifugation at 100,000g for 2 h at 4 °C. Afterwards, the supernatant was discarded. Exosomes used for RNA extraction were resuspended in 500 μl of Trizol; exosomes used for protein extraction were resuspended in 250 μl of lysis buffer (8 M urea, 2.5% SDS, 5 μg ml^{-1} leupeptin, 1 μg ml^{-1} pepstatin and 1 mM phenylmethylsulphonyl fluoride). Exosomes used for flow cytometry analysis (FACS), TEM (see sections below) and immunogold staining were resuspended in 100 μl PBS. Ten microlitres of these exosomes sample were used for NanoSight LM10 (NanoSight Ltd) analysis after dilution 1:100 in PBS.

Exosomes isolation from human serum samples. As previously described, 250 μl of cell-free serum samples were thawed on ice[6]. Serum was diluted in 11 ml PBS and filtered through a 0.2-μm pore filter. Afterwards, the samples were ultracentrifuged at 150,000g overnight at 4 °C. Next, the exosomes pellet was washed in 11 ml PBS followed by a second step of ultracentrifugation at 150,000g at 4 °C for 2 h. The supernatant was discarded and pelleted exosomes were resuspended in 500 μl of Trizol for RNA analyses; or in 250 μl of lysis buffer (8 M urea, 2.5% SDS, 5 μg ml^{-1} leupeptin, 1 μg ml^{-1} pepstatin and 1 mM phenylmethylsulphonyl fluoride) for protein analyses. Exosomes used for flow cytometry analysis (FACS), TEM (see sections below) and immunogold staining were resuspended in 100 μl PBS. Ten microlitres of this exosomes sample were used for NanoSight LM10 (NanoSight Ltd) analysis after Nano dilution 1:100 in PBS.

Immunogold labelling and electron microscopy. Fixed specimens at an optimal concentration were placed onto a 400-mesh carbon/formvar coated grids and allowed to absorb to the formvar for a minimum of 1 min. For immunogold staining the grids were placed into a blocking buffer for a block/permeabilization step for 1 h. Without rinsing, the grids were immediately placed into the primary antibody at the appropriate dilution overnight at 4 °C (1:300 anti-CD9 ab92726, Abcam and anti-GPC1 PIPA528055, Thermo Scientific). As controls, some of the

grids were not exposed to the primary antibody. The next day, all the grids were rinsed with PBS then floated on drops of the appropriate secondary antibody attached with 10-nm gold particles (AURION) for 2 h at room temperature. Grids were rinsed with PBS and were placed in 2.5% glutaraldehyde in 0.1 M phosphate buffer for 15 min. After rinsing in PBS and distilled water the grids were allowed to dry and stained for contrast using uranyl acetate. The samples were viewed with a Tecnai Bio Twin transmission electron microscope (FEI) and images were taken with an AMT CCD Camera (Advanced Microscopy Techniques).

Sucrose gradient. Sucrose density gradients were performed to purify exosomes. Exosomes were resuspended in 2 ml of HEPES/sucrose stock solution (2.5 M sucrose, 20 mM HEPES/NaOH solution, pH 7.4). The exosomes suspension was overlaid with a linear sucrose gradient (2.0–0.25 M sucrose, 20 mM HEPES/NaOH, pH 7.4) in a SW41 tube (Beckman). The gradients were ultracentrifuged for 16 h at $210,000g$ at 4 °C. Gradient fractions of 1 ml were collected from top to bottom and densities of each fractions were evaluated using a refractometer. Next, the exosomes pellets were washed in PBS followed by a second step of ultracentrifugation at $150,000g$ at 4 °C for 2 h. Exosomes pellets were resuspended in Laemmli buffer and/or PBS for further immunoblot and flow cytometry analysis.

Flow cytometry analysis of exosomes-bound beads. Exosomes were attached to 4-μm aldehyde/sulphate latex beads (Invitrogen) by mixing 30 μg exosomes in a 10 μl volume of beads for 15 min at room temperature with continuous rotation. This suspension was diluted to 1 ml with PBS and left for 30 min rotating at room temperature. The reaction was stopped with 100 mM glycine and 2% BSA in PBS and left rotating for 30 min at room temperature. Exosomes-bound beads were washed once in 2% BSA in PBS and centrifuged for 1 min at $14,800g$, blocked with 10% BSA with rotation at room temperature for 30 min, washed a second time in 2% BSA and centrifuged for 1 min at $14,800g$, and incubated with anti-GPC1 (PIPA528055, Thermo-Scientific, 3 μl of antibody in 20 μl of 2% BSA) during 30 min rotation at 4 °C. Beads were centrifuged for 1 min at $14,800g$, the supernatant was discarded and beads were washed in 2% BSA and centrifuged for 1 min at $14,800g$. Alexa-488 or Alexa-594-tagged secondary antibodies (Life Technologies, 3 μl of antibody in 20 μl of 2% BSA) were used during 30 min with rotation at 4 °C. Secondary antibody incubation alone was used as control and to gate the beads with GPC1$^+$-bound exosomes. The percentage of positive beads was calculated relative to the total number of beads analysed per sample (100,000 events). This percentage was therein referred to as the percentage of beads with GPC1$^+$ exosomes.

UPLC–MS. Exosomes were mixed with 200 μl of methanol spiked with the internal standard tryptophan-d5. After brief vortex mixing, the samples were incubated for 1 h at −20 °C. After centrifugation at $16,000g$ for 15 min at 4 °C, 190 μl of the supernatants was collected and the solvent removed. The dried extracts were then reconstituted in 15 μl of methanol, of which 10 μl were transferred to microtubes and derivatized. Chromatographic separation and mass spectrometric detection conditions are described in Supplementary Table 2. The mass range, 50–1,000 m/z, was calibrated with cluster ions of sodium formate. An appropriate test mixture of standard compounds was analysed before and after the entire set of randomized duplicated sample injections, to examine the retention time stability and sensitivity of the LC–MS system throughout the course of the run. Data were processed using the TargetLynx application manager for MassLynx 4.1 software (Waters Corp.). A set of predefined retention time, mass-to-charge ratio pairs (RT-m/z), corresponding to metabolites included in the analysis are fed into the program. Associated extracted ion chromatograms (mass tolerance window = 0.05 Da) are then peak-detected and noise-reduced in both the LC and MS domains such that only true metabolite related features are processed by the software. A list of chromatographic peak areas is then generated for each sample injection, using the RT-m/z data pairs (retention time tolerance = 6 s) as identifiers. Normalization factors were calculated for each metabolite by dividing their intensities in each sample by the recorded intensity of the internal standard in that same sample. Visualization of disjoint and overlapping protein data sets was carried out by drawing a VennDiagram of the 5 protein data sets using an R package[45].

CA19-9 human and GPC1 ELISAs. Serum CA19-9 and GPC1 protein levels in patients with pancreatic cancer, pancreatic cancer precursor lesion, or benign pancreatic disease, and in healthy donors were assessed using the Cancer Antigen CA19-9 Human ELISA Kit (Abcam, ab108642) and the GPC1 Human ELISA kit (ABIN840422), according to the manufacturer's directions.

Western blot analyses. Cells were lysed in RIPA buffer containing 5 μg ml^{-1} leupeptin, 1 μg ml^{-1} pepstatin and 1 mM phenylmethylsulphonyl fluoride. Exosomes were lysed in 8 M urea, 2.5% SDS containing 5 μg ml^{-1} leupeptin, 1 μg ml^{-1} pepstatin and 1 mM phenylmethylsulphonyl fluoride. Sample loading was normalized according to Bradford relative protein quantification and proteins separated following an electrophoretic gradient across polyacrylamide gels. Wet electrophoretic transfer was used to transfer the proteins in the gel

onto PVDF membranes (ImmobilonP). The protein blot was blocked for 1 h at room temperature with 5% non-fat dry milk in PBS/0,05% Tween and incubated overnight at 4 °C with the following primary antibodies: 1:300 anti-GPC1, PIPA528055 (Thermo-Scientific); 1:300 anti-β-actin A3854 (Sigma-Aldrich); 1:300 anti-CD81 sc-166029 (Santa-Cruz); 1:300 anti-flottilin1 sc-25506 (Santa-Cruz). Afterwards, horseradish peroxidase (HRP)-conjugated secondary antibodies were incubated for 1 h at room temperature. Washes after antibody incubations were done on an orbital shaker, four times at 10-min intervals, with PBS 0.05% Tween20. Blots were developed with chemiluminescent reagents from Pierce.

RNA extraction of cells and exosomes. RNA of cells and exosomes was isolated using Trizol Plus RNA purification kit (Life Technologies, 12183555) according to manufactures protocol. RNA was quantified using a Nanodrop ND-1000 (Thermo Fischer Scientific).

qRT–PCR. Quantitative reverse transcriptase PCR (qRT–PCR) was performed on DNase-treated RNA using the SuperScript III Platinum One-Step Quantitative RT–PCR System (11732-088, Invitrogen) according to the manufacturer's directions on a 7300 Sequence Detector System (Applied Biosystems). 150 ng of RNA extracted from 2.5×10^8 exosomes was used as qPCR input. Primers for $KRAS^{G12D}$ and $KRAS^{G12V}$ mRNA (both Sigma-Aldrich) were designed as reported previously[46]. In brief, the altered base of the $KRAS^{G12D}$ and $KRAS^{G12V}$ mutation was kept at the 3′ end of the forward primer. An additional base mutation was included two positions before the $KRAS$ mutation to increase the specificity of the amplification of the mutant $KRAS$ allele. Forward primer sequence for $KRAS^{G12D}$ mRNA: F-5′-ACTTGTGGTAGTTGGAGC*A GA*-3′ (italicized bases denote mutations corresponding to the $KRAS$ mutant). Forward primer sequences for $KRAS^{G12V}$ mRNA: F-5′-ACTTGTGGTAGTTGG AGC*A GT*-3′. Forward primer sequences for $KRAS$ wild-type mRNA: F-5′-AC TTGTGGTAGTTGGAGCTGG-3′. Reverse primer for all $KRAS$ mRNAs: R-5′-TTGGATCATATTCGTCCACAA-3′. $GPC1$ mRNA primer pairs (PPH06045A) and 18S mRNA primer pairs (QF00530467) were purchased as ready specific primer pairs from Qiagen. Threshold cycle[47] (C_t) the fractional cycle number at which the amount of amplified target reached a fixed threshold, was determined and expression was measured using the $2^{-\Delta C_t}$ formula, as previously reported[48].

DNA extraction from human primary pancreatic cancer tumours and crExos. Immediately after resection, pancreatic tumour samples were snap-frozen in liquid nitrogen and stored at −80 °C until further processing. A 10-μm reference section of each sample was cut and stained with haematoxylin and eosin by standard methods to evaluate the proportion of tumour tissue and adjacent tumour stroma. Samples with a tumour stroma proportion >30% were excluded into this study. DNA isolation was performed using a commercial DNA Extraction Kit (DNeasy Blood & Tissue Kit, 69506, Qiagen) according to the manufacturer's protocol. The amount of DNA from tumour samples was quantified using a Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific).

PCR and Sanger sequencing. PCR was performed in a 25-μl reaction tube consisting of 10 μl template DNA, 1 μM of each primer, 2.5 mM dNTP, 2.5 μl 10× PCR buffer, 25 mM Mg solution, 0.5 μl H$_2$O and 2.5 μl Taq polymerase. Amplification was carried out in a T100 ThermoCycler (Bio-Rad) under the following conditions: 94 °C for 1 min, 2 cycles of 94 °C for 10 s, 67 °C for 30 s, 70 °C for 30 s; 2 cycles of 94 °C for 10 s, 64 °C for 30 s, 70 °C for 30 s; 2 cycles of 94 °C for 10 s, 61 °C for 30 s, 70 °C for 30 s; 35 cycles of 94 °C for 10 s, 59 °C for 30 s, 70 °C for 30 s; endless 4 °C. $KRAS$ amplicon were generated using the following primers: forward 5′-AAGGCCTGCTGAAAATGACTG-3′, 5′-AGAATGGTCC TGCACCAGTAA-3′. PCR products were purified using the QIAquick PCR purification kit (Qiagen). Subsequently, sequencing reaction was performed using BigDye terminator kit (v3.1, Life Technologies) according to the manufacturer's instructions. Sequencing products were separated on an ABI 3730 automated sequencer (Life Technologies). KRAS mutation status was evaluated using Finch TV (Geospiza, Inc.).

MRI imaging. MRI studies were conducted using a 7T small animal MR system, the Biospec USR70/30 (Bruker Biospin MRI) is based on an actively shielded 7T magnet with a 30-cm bore and cryo-refrigeration. The system is equipped with 6 cm inner-diameter gradients that deliver a maximum gradient field of 950 mT m^{-1}. A 3.5 cm inner-diameter linear birdcage coil transmits and receives the MR signal. For image acquisition, T2-weighted, respiratory gated, multi-slice imaging will be performed with respiration held to under 25 breaths per minute to minimize motion artefacts in the abdomen. For mice where fat signal might mask the T2 weighted image the fat-suppression pulse module will be used. Acquisition parameters were minimally modified from ref. 49. The rapid acquisition with relaxation enhancement (RARE) T2-weighted pulse sequence was modified to include an effective Te (time of echo) of 56 ms with a total TR (time repetition) of 2,265 ms. Between 18 and 20 coronal slices were acquired per mouse with a slice thickness of 0.75 mm and slice spacing of 1 mm. In plane, pixel sizes of 0.156 mm × 0.156 mm with a matrix size of 256 × 192 (40 mm × 30 mm FOV)

was chosen to minimize in plane partial volume effects, maintain a FOV sufficient to cover the abdomen, while also providing sufficient throughput for the experiment.

To measure tumour burden, the region of suspected lesions are drawn blinded on each slice after images intensities were normalized. The volume is calculated by addition of delineated region of interest in $mm^2 \times 1\,mm$ slice distance.

Statistical analysis. The GraphPad Prism version 6.0 (GraphPad Software) and MedCalc statistical software version 13.0 (MedCalc Software bvba) were used for all calculations. Unpaired Student's *t*-test was applied to calculate expression differences of the qPCR results (ΔC_t values). ANOVA tests were performed to calculate differences of multiple serum factors in murine and human serum samples. As a post-hoc test, a Tukey–Kramer test was applied for pairwise comparison of subgroups when the ANOVA test was positive in case of equal variance. Tamhane T_2 test was applied for pairwise comparison of subgroups when the ANOVA test was positive in case of unequal variances. A paired two-tailed Student's *t*-test was applied to assay differences in the percentage of beads with GPC1[+] crExos and CA19-9 in the longitudinal cohort between pre-operative and postoperative blood samples. ROC curves were used to determine the sensitivity, specificity, positive and negative predictive values and to compare AUCs of serum factors using the Delong method[50]. The cut-off value was determined using the Youden index. Univariate analysis using the log-rank test was conducted to visualize (Kaplan–Meier curves) and assess disease-specific survival (time from diagnosis to cancer-related death or last follow-up) in the longitudinal cohort of patients with pancreatic cancer. A multivariate analysis using the Cox proportional hazards regression model was performed to evaluate the effect of a decrease of the percentage of beads with GPC1[+] crExos in addition to age (continuous variable), AJCC tumour stage, and tumour grade (G) and CA19-9 levels (U ml[−1]). Correlation analysis between murine tumour burden and percentage beads with GPC1[+] crExos was performed using the Spearman correlation test. Figures were prepared using GraphPad Prism and MedCalc statistical software version 13.0. All presented *P* values are two-sided and $P < 0.05$ was considered to be statistically significant.

43. Brand, R. E. *et al.* Serum biomarker panels for the detection of pancreatic cancer. *Clin. Cancer Res.* **17,** 805–816 (2011).
44. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4,** 437–450 (2003).
45. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12,** 35 (2011).
46. Rachagani, S. *et al.* Activated Kras[G12D] is associated with invasion and metastasis of pancreatic cancer cells through inhibition of E-cadherin. *Br. J. Cancer* **104,** 1038–1048 (2011).
47. Rothstein, D. M. *et al.* Targeting signal 1 through CD45RB synergizes with CD40 ligand blockade and promotes long term engraftment and tolerance in stringent transplant models. *J. Immunol.* **166,** 322–329 (2001).
48. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25,** 402–408 (2001).
49. Schmid, A., Braumuller, H., Wehrl, H. F., Rocken, M. & Pichler, B. J. Non-invasive monitoring of pancreatic tumor progression in the RIP1-Tag2 mouse by magnetic resonance imaging. *Mol. Imaging Biol.* **15,** 186–193 (2013).
50. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44,** 837–845 (1988).
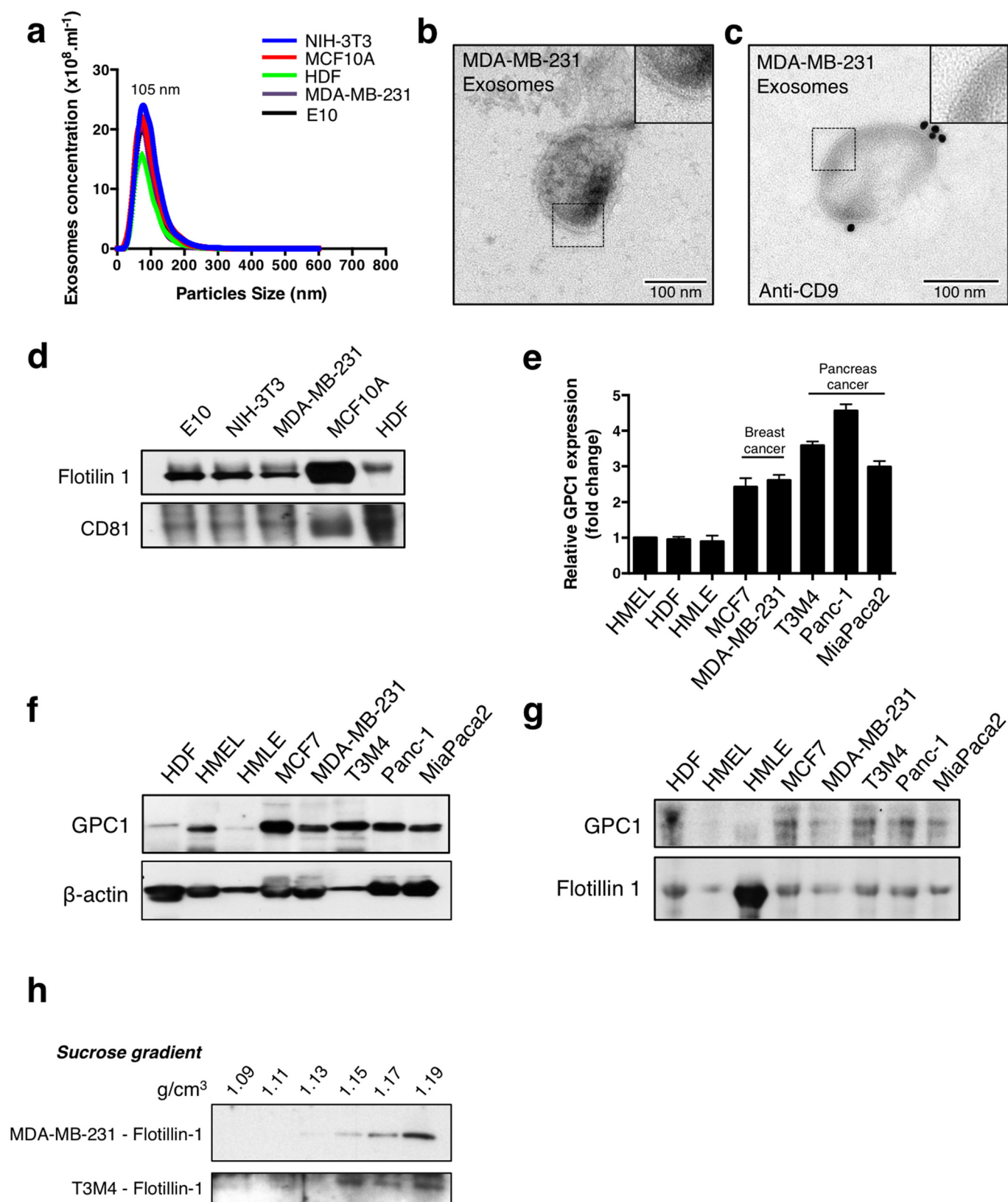
Extended Data Figure 1 | Exosome isolation. a, Exosome concentration and size distribution by NanoSight analysis of culture supernatant from NIH/3T3, MCF10A, HDF, MDA-MB-231 and E10 cells. Size mode: 105 nm (3 technical replicates). **b**, TEM micrograph of MDA-MB-231-derived exosomes. Top right image shows a digitally zoomed inset. **c**, TEM micrograph of MDA-MB-231-derived exosomes following immunogold labelling for CD9. Gold particles are depicted as black dots. Top right image shows a digitally zoomed inset. **d**, Immunoblot of flotillin1 and CD81 in exosomal proteins extracted from culture supernatant of E10, NIH/3T3, MDA-MB-231, MCF10A and HDF cells. **e**, qRT–PCR measurement of *GPC1* mRNA levels in HMEL, HDF, HMLE, MCF7, MDA-MB-231, T3M4, Panc-1 and MIA Paca2 cells. Results are mean ± s.d.; $n = 3$, 3 biological replicates, with 3 technical replicates each. **f**, Immunoblot of GPC1 in HMEL, HDF, HMLE, MCF7, MDA-MB-231, T3M4, Panc-1 and MIA Paca2 cell lysates (top). β-actin was used as a loading control (bottom). **g**, Immunoblot of GPC1 in exosomal protein lysates derived from the culture supernatant of 3 non-tumorigenic cell lines (HDF, HMEL and HMLE) and 5 tumorigenic cell lines (MCF7, MDA-MB-231, T3M4, Panc-1 and MIA Paca2) (top). Flotillin1 was used as loading control (bottom). **h**, Immunoblot of flotillin1 in exosomal protein lysates from the culture supernatant of MDA-MB-231 and T3M4 following sucrose gradient purification. The protein content is assayed in each of the density layers listed.

a Longitudinal Mouse Study

b

c
r= 0.98
P= 0.004

d
— MDA-MB-231-CD63GFP
— MDA-MB-231-CD63GFP 488

e

Circulating exosomes
MDA-MB-231-CD63GFP

Circulating exosomes
MDA-MB-231-CD63GFP - 488

f

MDA-MB-231 Exos

MDA-MB-231-CD63GFP Exos

crExos from mice with MDA-MB-231 CD63GFP tumors
2nd antibody only

crExos from mice with MDA-MB-231 CD63GFP tumors

3.2%

97.3%
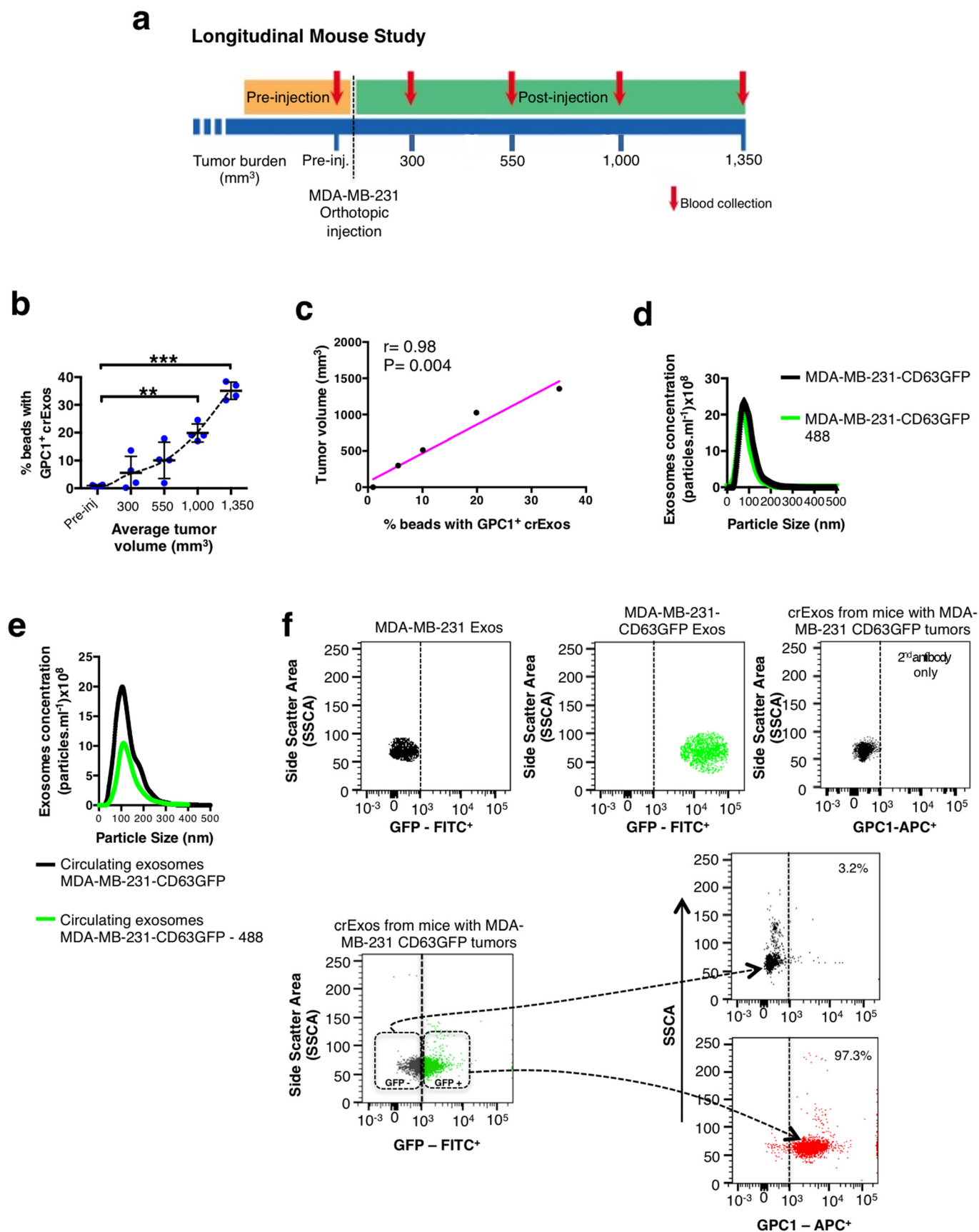
Extended Data Figure 2 | GPC1$^+$ crExos are derived from cancer cells in tumour-bearing mice. a, Longitudinal blood collection; nude mice with orthotopic MDA-MB-231 tumours ($n = 4$ mice). **b**, Percentage of beads with GPC1$^+$ crExos plotted against average tumour volume ($n = 4$ mice, each sample analysed in technical triplicates for GPC1). ANOVA, post-hoc Tamhane T$_2$, **$P < 0.01$, ***$P < 0.001$. Data are mean $\pm$ s.d. **c**, Correlation between tumour volume and the percentage of beads with GPC1$^+$ crExos (Pearson correlation test). **d**, NanoSight of exosomes from MDA-MB-231-CD63–GFP cells. Black: all exosomes; green: CD63–GFP$^+$ exosomes ($n = 3$ technical replicates). **e**, NanoSight of crExos from mice with a MDA-MB-231-CD63–GFP orthotopic tumour. Black: all exosomes; green: CD63–GFP$^+$ exosomes ($n = 3$ technical replicates). **f**, FACS analysis of beads with exosomes from cultured MDA-MB-231 (top left) and MDA-MB-231-CD63–GFP (top middle) cells, and crExos of mice with MDA-MB-231-CD63–GFP orthotopic tumours (bottom left). Staining of CD63–GFP$^+$ (cancer-cell-derived) and CD63–GFP$^-$ (host-derived) crExos for GPC1 (allophycocyanin (APC)$^+$ bottom right; $n = 3$ biological replicates and 3 technical replicates). The percentage of positive beads is listed. Negative control: secondary antibody alone (top right).
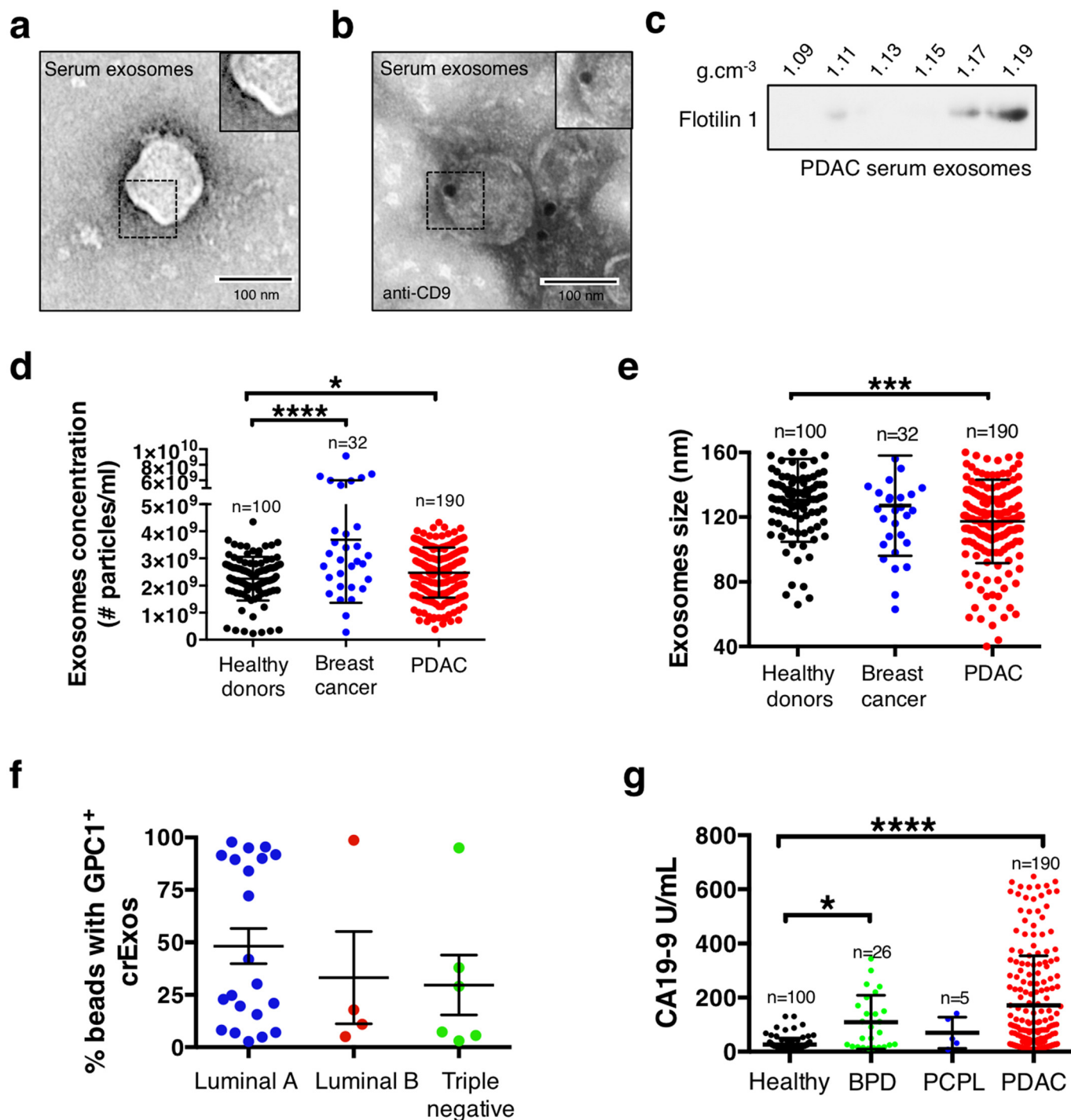
Extended Data Figure 3 | NanoSight analysis in human serum samples.
a, TEM micrograph of serum-derived exosomes from a cancer patient. Top right image shows a digitally zoomed inset. **b**, TEM micrograph of serum-derived exosomes from a cancer patient after immunogold labelling for CD9. Gold particles are depicted as black dots. Top right image shows a digitally zoomed inset. **c**, Immunoblot of flotillin1 of exosomal protein lysates from serum of cancer patient following exosome purification by a sucrose gradient. The protein content is assayed in each of the density layers listed. **d**, Exosome concentration by NanoSight analysis showing the number of exosomes per millilitre of serum derived from healthy donors ($n = 100$), breast cancer patients ($n = 32$) and patients with PDAC ($n = 190$) (ANOVA, post-hoc Tamhane T_2, *$P < 0.05$, ****$P < 0.0001$; 3 technical replicates). **e**, Exosomes size distribution by NanoSight analysis showing the mode size of exosomes in 1 ml of serum derived from healthy donors ($n = 100$), breast cancer patients ($n = 32$) and patients with PDAC ($n = 190$) (ANOVA, post-hoc Tukey–Kramer test, ***$P < 0.001$; 3 technical replicates). **f**, Scatter dot plots depicting the percentage of beads with GPC1[+]-bound exosomes purified from the serum of breast cancer patients. The patients are subdivided into three subtypes: luminal A, luminal B and triple-negative breast cancer. **g**, Scatter plots depicting the serum CA19-9 concentration (U ml^{-1}), evaluated by ELISA, in healthy donors ($n = 100$), patients with BPD ($n = 26$), PCPL ($n = 5$) and PDAC ($n = 190$). Discovery cohort, ANOVA, post-hoc Tamhane T_2, *$P < 0.05$; ****$P < 0.0001$; 3 technical replicates. Data are mean ± s.d.

a

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI2
Beads with GPC1+ crExos (%)	1	0.988 - 1.000	>7.6	100	98.1 - 100.0	100	97.1 - 100.0
CA 19-9 (U/ml)	0.739	0.687 - 0.787	>26.3063	76.84	70.2 - 82.6	64.29	55.3 - 72.6
Exosomes Concentration (^10E09)	0.57	0.513 - 0.625	>32.8	25.79	19.7 - 32.6	92.06	85.9 - 96.1
Exosome Size (nm)	0.676	0.621 - 0.727	≤122	63.16	55.9 - 70.0	70.63	61.9 - 78.4

b

CIS Stage I GPC1+ · CA 19-9 · Exosomes concentration · Exosomes size

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI
GPC-1-positive exosomes (%)	1	0.972 - 1.000	>7.6	100	47.8 - 100.0	100	97.1 - 100.0
CA 19-9 (U/ml)	0.735	0.651 - 0.808	>30.8435	80	28.4 - 99.5	66.67	57.7 - 74.8
Exosomes Concentration (^10E09)	0.581	0.492 - 0.667	≤23.75E08	60	14.7 - 94.7	43.65	34.8 - 52.8
Exosomes Size (nm)	0.663	0.576 - 0.744	≤107	60	14.7 - 94.7	88.1	81.1 - 93.2

c

PDAC Stage IIa GPC1+ · CA 19-9 · Exosomes concentration · Exosomes size

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI
GPC-1-positive exosomes (%)	1	0.975 - 1.000	>7.6	100	81.5 - 100.0	100	97.1 - 100.0
CA 19-9 (U/ml)	0.668	0.585 - 0.744	>26.3063	66.67	41.0 - 86.7	64.29	55.3 - 72.6
Exosomes Concentration (^10E09)	0.648	0.564 - 0.726	>28.1E08	55.56	30.8 - 78.5	78.57	70.4 - 85.4
Exosomes Size (nm)	0.7	0.619 - 0.774	≤124	77.78	52.4 - 93.6	66.67	57.7 - 74.8

d

PDAC Stage IIb GPC1+ · CA 19-9 · Exosomes concentration · Exosomes size

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI
GPC-1-positive exosomes (%)	1	0.985 - 1.000	>7.6	100	96.9 - 100.0	100	97.1 - 100.0
CA 19-9 (U/ml)	0.74	0.680 - 0.794	>25.3562	79.49	71.0 - 86.4	63.49	54.4 - 71.9
Exosomes Concentration (^10E09)	0.559	0.494 - 0.622	>31.7E08	27.35	19.5 - 36.4	89.68	83.0 - 94.4
Exosomes Size (nm)	0.692	0.630 - 0.749	≤122	66.67	57.4 - 75.1	70.63	61.9 - 78.4

e

PDAC Stage III GPC1+ · CA 19-9 · Exosomes concentration · Exosomes size

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI
GPC-1-positive exosomes (%)	1	0.973 - 1.000	>7.6	100	71.5 - 100.0	100	97.1 - 100.0
CA 19-9 (U/ml)	0.729	0.646 - 0.801	>36.1015	72.73	39.0 - 94.0	71.43	62.7 - 79.1
Exosomes Concentration (^10E09)	0.566	0.478 - 0.650	>32.8E08	36.36	10.9 - 69.2	92.06	85.9 - 96.1
Exosomes Size (nm)	0.776	0.697 - 0.842	≤132	100	71.5 - 100.0	49.21	40.2 - 58.3

f

PDAC Stage IV GPC1+ · CA 19-9 · Exosomes concentration · Exosomes size

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI
GPC-1-positive exosomes (%)	1	0.978 - 1.000	>7.6	100	91.4 - 100.0	100	97.1 - 100.0
CA 19-9 (U/ml)	0.788	0.718 - 0.848	>61.2284	75.61	59.7 - 87.6	78.57	70.4 - 85.4
Exosomes Concentration (^10E09)	0.569	0.490 - 0.645	>26.5E08	51.22	35.1 - 67.1	70.63	61.9 - 78.4
Exosomes Size (nm)	0.604	0.525 - 0.678	≤122	56.1	39.7 - 71.5	70.63	61.9 - 78.4

g

Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI2
Beads with GPC1+ crExos (%)	1	0.956 - 1.000	>12	100	93.6 - 100.0	100	86.8 - 100.0
Exosomes Concentration (^10E09)	0.714	0.603 - 0.808	>22.2E08	51.79	38.0 - 65.3	84.62	65.1 - 95.6
Exosome Size (nm)	0.648	0.534 - 0.750	≤121	62.5	48.5 - 75.1	65.38	44.3 - 82.8

Extended Data Figure 4 | Tumour-stage-specific analysis. a, Table associated with ROC curve analysis depicted in Fig. 1f. **b–f**, ROC curve analysis for the percentage of GPC1$^+$ crExos (red line), CA19-9 serum levels (blue scattered line), exosome concentration (black line) and exosome size (scattered black line) in patients with carcinoma *in situ* (CIS) or stage I pancreatic cancer ($n = 5$) (**a**), stage IIa pancreatic cancer ($n = 18$) (**b**), stage IIb pancreatic cancer ($n = 117$) (**c**), stage III pancreatic cancer ($n = 11$) (**d**), and stage IV pancreatic cancer ($n = 41$) (**e**), compared to healthy donors ($n = 100$) and patients with a benign pancreatic disease ($n = 26$), total $n = 126$. **g**, Table associated with ROC curve analysis depicted in Fig. 1h. CI, confidence interval.

a
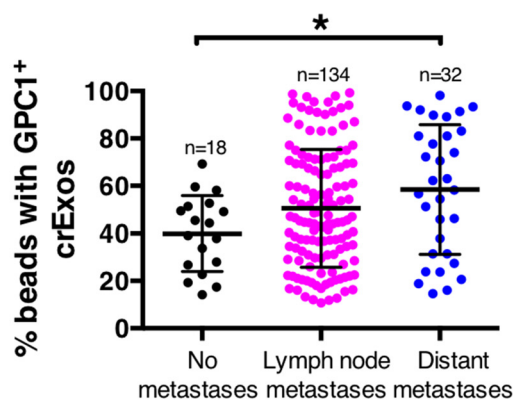
b
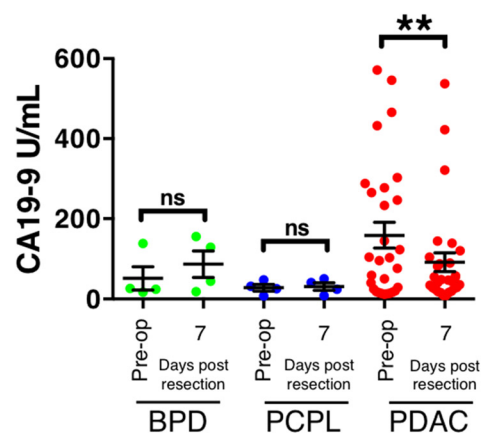
c

Multivariate analysis – overall survival

Parameter	Hazard ratio	95% CI	P - Value
GPC1 drop between day 0 and day 7	5.511	1.697-17.892	0.005
Age	0.96	0.898-1.026	0.227
AJCC stage	1.203	0.429-3.374	0.726
Tumor grade	1.024	1.004-1.044	0.018
CA 19-9 drop between day 0 and day 7	2.453	0.885-6.796	0.084

d

Multivariate analysis – disease specific survival

Parameter	Hazard ratio	95% CI	P - Value
GPC1 drop between day 0 and day 7	5.353	1.651-17.358	0.005
Age	0.962	0.899-1.028	0.254
AJCC stage	1.177	0.428-3.237	0.752
Tumor grade	1.016	0.992-1.041	0.197
CA 19-9 drop between day 0 and day 7	2.138	0.762-5.993	0.149

e
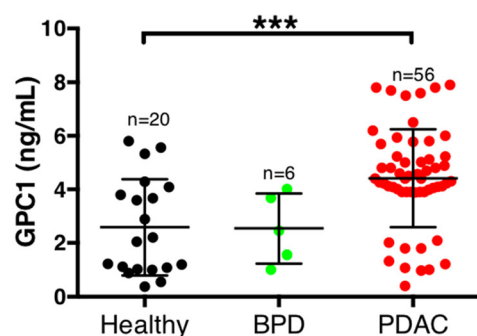
f
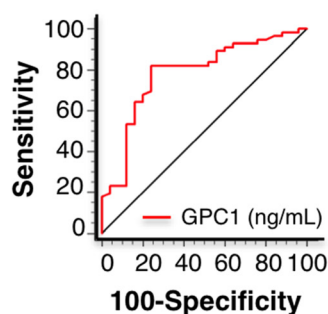
Parameter	AUC	CI	Cut-off value	Sensitivity	95% CI	Specificity	95% CI
GPC-1 ELISA (ng/ml)	0.781	0.675 - 0.865	>3.8	82.14	69.6 - 91.1	76	54.9 - 90.6

Extended Data Figure 5 | Longitudinal human study. a, Scatter plots of the percentage of beads with $GPC1^+$ crExos by flow cytometry in patients with pancreatic cancer. Patients are divided based on metastatic disease (non-metastatic lesions, lymph node metastases and distant metastases) (ANOVA, post-hoc Tukey–Kramer test, $*P < 0.05$; 3 technical replicates). **b**, Scatter plots depicting serum CA19-9 levels (U ml^{-1}) in patients with BPD ($n = 4$), PCPL ($n = 4$) and PDAC ($n = 29$) on the pre-operative day and post-operative day 7 in patients (paired two-tailed Student's t-test, $**P < 0.01$; 3 technical replicates). **c**, **d**, Multivariate analysis (Cox proportional hazards regression model) of prognostic parameters for overall (**c**) and disease-specific (**d**) survival of patients with pancreatic cancer in the longitudinal cohort ($n = 29$). **e**, Scatter plots depicting serum GPC1 (ng ml^{-1}) levels by ELISA in patients with BPD ($n = 6$), PDAC ($n = 56$) and healthy controls ($n = 20$) (ANOVA, post-hoc Tukey–Kramer test, $**** P < 0.0001$; 3 technical replicates). **f**, ROC curve for circulating GPC1 protein (red line) in patients with pancreatic cancer ($n = 56$) versus healthy donors ($n = 20$) and patients with a benign pancreatic disease ($n = 26$), total $n = 6$.

a Ptf1a-Cre; Kras-LSL^{G12D}; TGFRβII$^{flox/flox}$ mice (PKT)

Extended Data Figure 6 | PDAC GEMM longitudinal study. a, Schematic diagram depicting the spontaneous development and progression of pancreatic cancer in PKT mice, and haematoxylin and eosin of the pancreas at the indicated time points showing healthy pancreas, and PanIN and PDAC lesions. Scale bars, 100 μm. **b, c**, Exosome size (**b**) and concentration (**c**) assayed by NanoSight analysis from the serum of PKT mice (E: experimental, red) and control mice (C: control, blue) at 4, 5, 6, 7 and 8 weeks of age (ANOVA, post-hoc Tukey–Kramer test, $*P < 0.05$; 3 technical replicates). **d**, Graph depicting the time-wise progression of tumour volume measured by MRI and the percentage of GPC1$^+$-bound crExo beads in individual PKT mice (blue: tumour volume, red: percentage of GPC1$^+$ crExos). **e**, Percentage of GPC1$^+$ crExo beads from control mice ($n = 3$) and mice with cerulein-induced acute pancreatitis ($n = 4$) (two-tailed Student's t-test, ns: not significant; 3 technical replicates). **f**, Results from ROC curves for the percentage of GPC1$^+$-bound crExo beads, exosome concentration and size in 4-, 5-, 6- and 7-week-old PKT mice ($n = 7$) versus control (including age-matched littermate healthy control ($n = 6$) and mice with induced acute pancreatitis ($n = 4$), $n = 10$)). Data are mean ± s.d.

a

PKT GEMM Cross Sectional Study

b
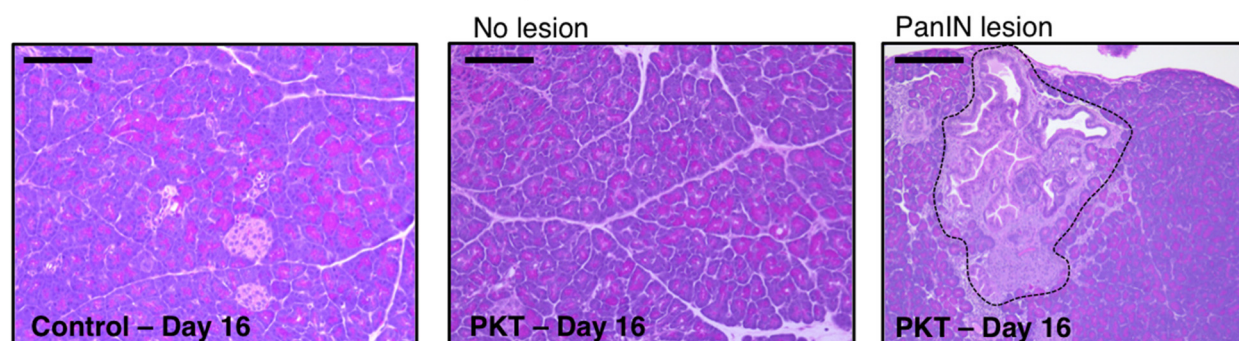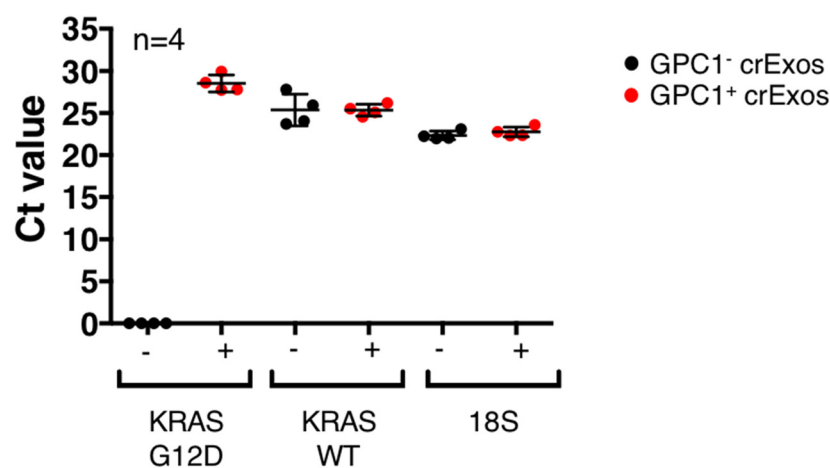
Extended Data Figure 7 | PDAC GEMM cross-sectional studies.
a, Representative micrographs of haematoxylin-and-eosin-stained pancreas from 16-day-old control mice (left) and PKT mice presenting with (right, encircled) and without (middle) PanIN lesions. Scale bars, 100 μm. **b**, C_t values following qPCR analyses for oncogenic $KRAS^{G12D}$, wild-type $KRAS$ and 18S internal control RNA from exosomes of 44–48-day-old PKT mice serum segregated using FACS for GPC1$^+$-bead-bound exosomes (red) and GPC1$^−$-bead-bound exosomes (blue). Data are mean $\pm$ s.d.

a

Healthy control exos – secondary only
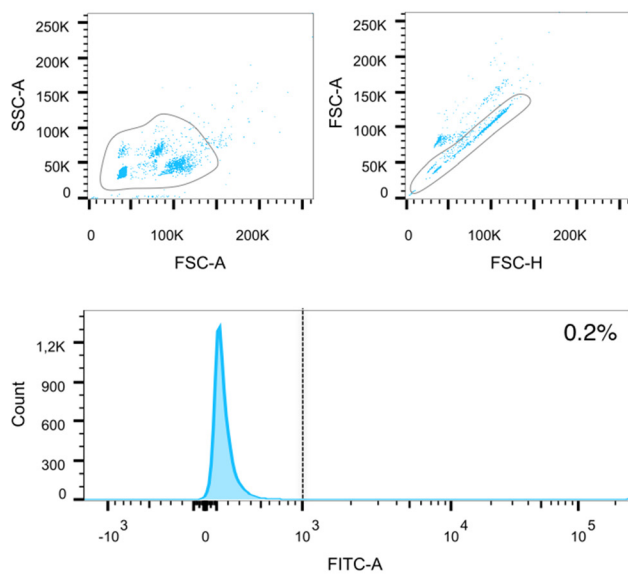
Healthy control exos – GPC1
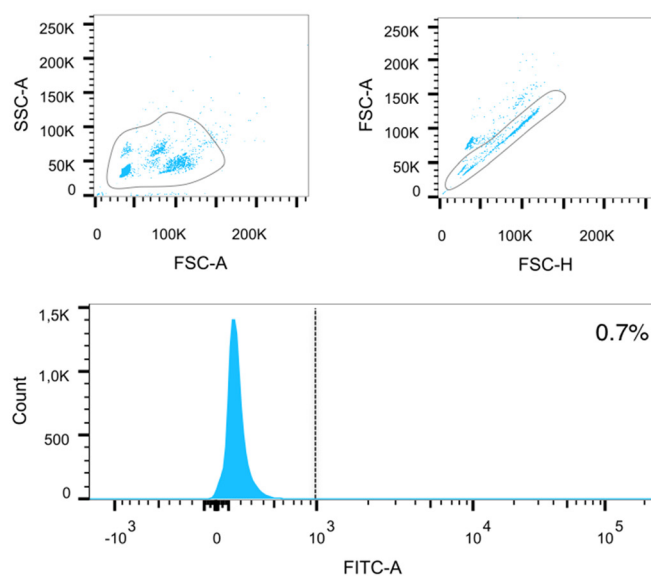
b

PDAC exos– secondary only
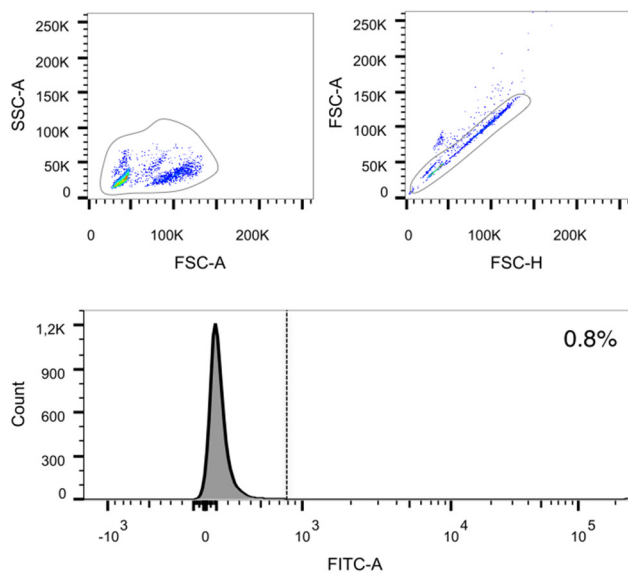
PDAC exos– GPC1
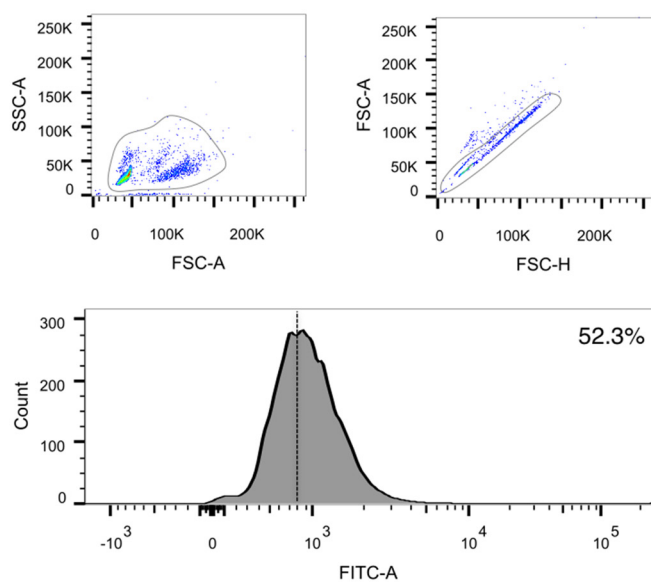
Extended Data Figure 8 | Raw scatter dot plot depicting flow cytometry analyses of beads with GPC1[+]-bound exosomes a, Scatter plots and histogram of flow cytometry analyses of serum exosomes on beads of a representative healthy control (left panels are secondary antibody only; right panels are GPC1 antibody and secondary antibody). **b**, Scatter plots and histogram of flow cytometry analysis of serum exosomes on beads of a representative pancreatic cancer sample (left panels are secondary antibody only; right panels are with GPC1 antibody and secondary antibody).

Extended Data Table 1 | The 48 proteins exclusive to MDA-MB-231 exosomes and histopathological findings and scoring in PKT mice in the cross-sectional study

a

Protein Name	Gene ID	Cellular Location
ATP-binding cassette sub-family A member 6	ABCA6	Transmembrane
Tetraspanin-4	TSPAN4	Transmembrane
SLIT and NTRK-like protein 4	SLITRK4	Transmembrane
Putative protocadherin beta-18	PCDHB18	Transmembrane
Myeloid cell surface antigen CD33	CD33	Transmembrane
Glypican-1	GPC1	Membrane anchored

Protein Name	Gene ID	Cellular Location
Histone H2A type 2-A	HIST1H2AA	Nucleus
Histone H2A type 1-A	HIST1H1AA	Nucleus
Histone H3.3	H3F3A	Nucleus
Histone H3.1	HIST1H3A	Nucleus
Zinc finger protein 37 homolog	ZFP37	Nucleus
Hypermethylated in cancer 2 protein	HIC2	Nucleus
Zinc finger protein 12	ZSCAN12	Nucleus

Protein Name	Gene ID	Cellular Location
Laminin subunit beta-1	LAMB1	Secreted
Tubulointerstitial nephritis antigen-like	TINAGL1	Secreted
Peroxiredoxin-4	PRDX4	Secreted
Collagen alpha-2(IV) chain	COL4A2	Secreted
Putative protein C3P1	C3P1	Secreted
Collagen alpha-1(II) chain	COL2A1	Secreted
Hemicentin-1	HMCN1	Secreted

Protein Name	Gene ID	Cellular Location
Junction plakoglobin	JUP	Cytoplasm
Tubulin beta-2B chain	TUBB2B	Cytoplasm
Endoribonuclease Dicer	DICER1	Cytoplasm
E3 ubiquitin-protein ligase TRIM71	TRIM71	Cytoplasm
Katanin p60 ATPase-containing subunit A-like 2	KATNAL2	Cytoplasm
Protein S100-A6	S100A6	Cytoplasm
5'-nucleotidase domain-containing protein 3	NT5DC3	Cytoplasm
Valine--tRNA ligase	VARS	Cytoplasm
Kazrin	KAZN	Cytoplasm
ELAV-like protein 4	ELAVL4	Cytoplasm
RING finger protein 166	RNF166	Cytoplasm
FERM and PDZ domain-containing protein 1	FRMPD1	Cytoplasm
78 kDa glucose-regulated protein	HSPA5	Cytoplasm
Trafficking protein particle complex subunit 6A	TRAPPC6A	Cytoplasm
Squalene monooxygenase	SQLE	Cytoplasm
Tumor susceptibility gene 101 protein	TSG101	Cytoplasm
Vacuolar protein sorting 28 homolog	VPS28	Cytoplasm
Prostaglandin F2 receptor negative regulator	PTGFRN	Cytoplasm
Isobutyryl-CoA dehydrogenase, mitochondrial	ACAD8	Cytoplasm
26S protease regulatory subunit 6B	PSMC4	Cytoplasm
Elongation factor 1-gamma	EEF1G	Cytoplasm
Titin	TTN	Cytoplasm
Tyrosine-protein phosphatase type 13	PTPN13	Cytoplasm
Triosephosphate isomerase	TPI1	Cytoplasm
Carboxypeptidase E	CPE	Cytoplasm

Protein Name	Gene ID	Cellular Location
Putative rhophilin-2-like protein	RHPN2P1	Not specified
Ankyrin repeat domain-containing protein 62	ANKRD62	Not specified
Tripartite motif-containing protein 42	TRIM42	Not specified

b

Mouse ID	Age (days)	Sex	Genotype	GPC1 (%)	Normal (% area)	Involved (% area)	PanIN1a	PanIN1b	PanIN2	PanIN3	DCA	Other (defined)	Pathological score
T187	16	M	Healthy control	0.6	100	0	0	0	0	0	0		
T188	16	M	Healthy control	1.1	100	0	0	0	0	0	0		
T157	20	M	Healthy control	0.2	100	0	0	0	0	0	0		
T158	20	M	Healthy control	0.2	100	0	0	0	0	0	0		
T163	20	M	Healthy control	0.2	100	0	0	0	0	0	0		
T164	20	M	Healthy control	0.4	100	0	0	0	0	0	0		
T183	16	M	PKT	8.4	98.6	1.4	P	0	0	0	0		1
T190	16	F	PKT	5.2	100.0	0.0	0	0	0	0	0	reactive ducts	0
T191	16	F	PKT	6.4	100.0	0.0	0	0	0	0	0	reactive ducts	0
T156	20	M	PKT	8.3	100.0	0.0	0	0	0	0	0		0
T159	20	F	PKT	10.8	95.1	4.9	P	P	P	0	0		3
T160	20	F	PKT	9.8	99.3	0.7	P	P	0	0	0		2
T165	20	F	PKT	9.2	96.1	3.9	P	P	0	0	0		2

P : present
0 : not detected

a, Listing of the 48 proteins exclusively detected in exosomes from MDA-MB-231 cells determined by UPLC–MS and comparative analyses of exosomes derived from NIH/3T3, MCF 10A, HDF, E10 and MDA-MB-231 cells. The proteins are grouped based on cellular location. **b**, The mouse ID, age, genotype and percentage of GPC1⁺ crExo beads of mice in the cross-sectional study are listed. A description of the histopathological findings and associated histological score is listed for PKT mice. Score 1: PanIN1a; score 2: PanIN1 a/b; score 3: PanIN2; score 4: PanIN3; score 5: ductal adenocarcinoma (DCA). P, present (lesions were detected). 0, no lesion detected.

Extended Data Table 2 | Demographics of patients and healthy participants and histological report of patients with chronic pancreatitis

a

		DISCOVERY COHORT		VALIDATION COHORT			
		No. of participants (n=321)	% of participants	No. of participants (n=82)	% of participants	No. of participants (n=32)	% of participants
		Pancreatic Cancer		**Pancreatic Cancer**		**Breast cancer**	
Total		190	59.19%	56	68.29%	32	100%
Sex							
	Men	104	54.74%	28	50.00%	0	0%
	Women	86	45.26%	28	50.00%	32	100%
Median Age (range)		66 (37 - 86)		70 (40 - 85)		57 (30 - 85)	
AJCC stage							
	0	n.a.	-	n.a.	-	2	6%
	I	2	1.05%	2	3.57%	12	38%
	II	n.a	-	n.a	-	17	53%
	IIa	19	10.00%	15	26.79%	n.a.	-
	IIb	117	61.58%	36	64.29%	n.a.	-
	III	11	5.79%	0	0.00%	1	3%
	IV	41	21.58%	3	5.36%	n.a.	-
Tumor grade							
	1	1	0.53%	1	1.79%	8	25%
	2	91	47.89%	35	62.5%	13	41%
	3	49	25.79%	19	33.93%	10	31%
	4	1	0.53%	0	0.00%	n.a.	-
	Unknown	48	25.26%	1	1.79%	1	3%
Tumor resected							
	Yes	152	80.00%	54	96.43%	32	100%
	No	38	20.00%	2	3.57%	0	0%
Neoadjuvant Radio-/Chemotherapy							
	Received	10	5.26%	0	0.00%	0	0%
	Not received	180	94.74%	56	100.00%	32	100%
		Benign Pancreatic disease (BPD)					
Total		26	8.15%				
Sex							
	Men	18	69.23%	3			
	Women	8	30.77%	3	50.00%		
					50.00%		
Median Age (range)		58.5 (31 - 77)		49 (43 - 56)			
Diagnosis							
	Chronic pancreatits	15	57.69%	6	100%		
	Autoimmune pancreatitis	3	11.54%	0	0.00%		
	Serous cystadenoma	8	30.77%	0	0.00%		
		Pancreatic cancer precursor lesion (PCPL)					
Total		5	1.55%	0	0.00%		
Sex							
	Men	2	40.00%	0	0.00%		
	Women	3	60.00%	0	0.00%		
Median Age (range)		65 (59 - 74)		-	-	-	
Neoplasms							
	IPMN	5	100.0%	0	0.00%		
		Healthy donors					
Total		100	31.15%	20	24.39%		

n.a. - non applicable

b

Patient No.	PanIN described	Histopathological report
1	No	Pancreatic tissue with chronic pancreatitis and extensive fibrosis and focal necrosis lipolytic and triptolytic areas.
2	No	Diffuse periductal lymphoplasmacytic infiltrates; severe periductal fibrosis and duct obstruction/disappearance; severe interlobular and acinar involvement; severe inflammatory storiform fibrosis and diffuse sclerosis; frequent venulitis and occasional arteritis; scattered and occasionally prominent lymphoid follicles
3	No	Chronic pancreatitis with periductal , inter- and intralobular fibrosis
4	No	Chronic pancreatitis and extensive fibrosis
5	No	Chronic recurrent and acute pancreatitis with plurifocal tryptolytic and lipolytic necrosis
6	No	Low-grade chronic pancreatitis with periductal fibrosis, in the present material no evidence of neoplastic events, no evidence of malignancy.
7	No	Chronic recurrent pancreatitis with some more pronounced fibrosis and intraductal calcifications. Chronic pancreatitis extends to the pancreas resection margin.
8	No	Chronic pancreatitis, cholangitis and papillitis with focally histomorphological aspect of an autoimmune, chronic sclerosing pancreatitis.
9	PanIN 1a	Pancreatic parenchyma (head of the pancreas) and peripancreatic fat and connective tissue with chronic recurrent pancreatitis with some areas fibrosis and abscesses. Pancreatic intraepithelial neoplasia (PanIN) Grade 1A.
10	No	Pancreatic parenchyma with perilobular fibrosis as well as dilated pancreatic ducts. In addition, peripancreatic fat and connective tissue with fibrosis. The finding represents a chronic pancreatitis.
11	PanIN 1a	Chronic-recurrent pancreatitis with pronounced fibrosis and dilated pancreatic ducts with focal inflammatory reactive epithelial cells. Older areas of organized necrosis and focal pancreatic intraepithelial neoplasia (PanIN) Grade 1A. Chronic pancreatitis also affects the pancreas resection margin.
12	No	Pancreatitis with focally accentuated, periductal, perilobular and intralobular fibrosis as well as smaller areas of organized fatty necrosis and presence of singel giant cells of foreign body type.
13	No	Chronic recurrent and acute pancreatitis with plurifocal tryptolytic and lipolytic necrosis with extensive destruction of the pancreatic parenchyma. Smaller secretion - and obliteration of pancreatic ducts with periductal fibrosis and localized squamous metaplasia. Peri- and interlobular fibrosis of the pancreatic parenchyma. Pancreatitis reaches the resection margin. At present, no neoplastic tissue, no evidence of malignancy.
14	No	Pancreatic tissue with some scarring chronic inflammation and chronic pancreatits. In the present material, no evidence of malignancy.
15	No	Tumor -free pancreatic tissue (surgical margins) with low periductal and interlobular fibrosis.

a, The group of IPMNs consist of 2 IPMN associated with a carcinoma *in situ*, 1 IPMN associated with an early adenocarcinoma of the pancreas (pT1), an IPMN with intermediate dysplasia and an IPMN with low-grade dysplasia. AJCC, American Joint Committee on Cancer. **b**, The histopathological report is listed for the 15 patients with chronic pancreatitis in the discovery cohort.

ARTICLE

Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS

Jan Krönke[1,2,3]*, Emma C. Fink[1,3]*, Paul W. Hollenbach[4], Kyle J. MacBeth[4], Slater N. Hurst[1], Namrata D. Udeshi[3], Philip P. Chamberlain[4], D. R. Mani[3], Hon Wah Man[4], Anita K. Gandhi[4], Tanya Svinkina[3], Rebekka K. Schneider[1], Marie McConkey[1], Marcus Järås[1], Elizabeth Griffiths[5], Meir Wetzler[5], Lars Bullinger[2], Brian E. Cathers[4], Steven A. Carr[3], Rajesh Chopra[4] & Benjamin L. Ebert[1,3]

Lenalidomide is a highly effective treatment for myelodysplastic syndrome (MDS) with deletion of chromosome 5q (del(5q)). Here, we demonstrate that lenalidomide induces the ubiquitination of casein kinase 1A1 (CK1α) by the E3 ubiquitin ligase CUL4–RBX1–DDB1–CRBN (known as CRL4^{CRBN}), resulting in CK1α degradation. CK1α is encoded by a gene within the common deleted region for del(5q) MDS and haploinsufficient expression sensitizes cells to lenalidomide therapy, providing a mechanistic basis for the therapeutic window of lenalidomide in del(5q) MDS. We found that mouse cells are resistant to lenalidomide but that changing a single amino acid in mouse Crbn to the corresponding human residue enables lenalidomide-dependent degradation of CK1α. We further demonstrate that minor side chain modifications in thalidomide and a novel analogue, CC-122, can modulate the spectrum of substrates targeted by CRL4^{CRBN}. These findings have implications for the clinical activity of lenalidomide and related compounds, and demonstrate the therapeutic potential of novel modulators of E3 ubiquitin ligases.

The immunomodulatory (IMiD) agents lenalidomide, thalidomide, and pomalidomide are the first drugs identified that promote the ubiquitination and degradation of specific substrates by an E3 ubiquitin ligase. These compounds bind CRBN[1], the substrate adaptor for the CRL4^{CRBN} E3 ubiquitin ligase, and modulate the substrate specificity of the enzyme. Each of these drugs induces degradation of two lymphoid transcription factors, IKZF1 and IKZF3, leading to clinical efficacy in multiple myeloma and increased interleukin-2 release from T cells[2–4]. However, it has not yet been determined whether degradation of distinct substrates may mediate additional activities and whether all IMiD compounds have the same substrate specificity.

Lenalidomide is also a highly effective treatment for myelodysplastic syndrome (MDS) with deletion of chromosome 5q (del(5q)), inducing cytogenetic remission in more than 50% of patients[5–7]. In vitro, lenalidomide selectively induces apoptosis of del(5q) MDS cells[8,9]. No biallelic deletions or loss of function mutations on the remaining allele have been detected in any of the genes in the del(5q) common deleted region, implying that MDS with del(5q) is a disease of haploinsufficiency[10,11]. We hypothesized that ubiquitination of a distinct CRBN substrate explains the efficacy of lenalidomide in del(5q) MDS.

Lenalidomide induces degradation of CK1α

In order to identify lenalidomide-regulated CRL4^{CRBN} substrates in myeloid cells, we applied stable isotope labelling of amino acids in cell culture (SILAC)-based quantitative mass spectrometry[12] to assess global changes in ubiquitination[13] and protein levels in the del(5q) myeloid cell line KG-1 (Fig. 1a, b, Extended Data Fig. 1, Extended Data Table 1a, b and Supplementary Table 1). Treatment with 1 μM lenalidomide significantly altered only seven K-ε-GG sites from five proteins out of 13,061 reproducible sites, demonstrating the highly specific effects of this drug. Moreover, lenalidomide significantly altered the protein abundance of 3 out of 5 differentially ubiquitinated

proteins. Consistent with previous studies, lenalidomide treatment decreased ubiquitination of CRBN ($P = 0.026$) and increased ubiquitination of IKZF1 ($P = 7.23 \times 10^{-6}$ and $P = 4.97 \times 10^{-4}$ for two distinct sites), with a reciprocal decrease in IKZF1 protein abundance ($P = 0.006$)[2,3].

In addition to IKZF1, we detected increased ubiquitination ($P = 0.04$) and decreased protein abundance ($P = 0.006$) of casein kinase 1A1 (CK1α) following treatment with 1 μM lenalidomide (Fig. 1a, b, Extended Data Fig. 1, Extended Data Table 1a, b and Supplementary Table 1). CK1α is encoded by the *CSNK1A1* gene, which is located in the del(5q) common deleted region, and is expressed at haploinsufficient levels in del(5q) MDS[10,14]. CK1α has been implicated in the biology of del(5q) MDS[15] and has been shown to be a therapeutic target in myeloid malignancies[16], and is therefore an attractive candidate for mediating the effects of lenalidomide in del(5q) MDS.

CK1α is a substrate of CRL4^{CRBN}

We sought to determine whether CK1α is a lenalidomide-dependent substrate of the CRL4^{CRBN} E3 ubiquitin ligase. We confirmed that lenalidomide treatment decreases CK1α protein levels in multiple human cell lines and in the bone marrow and peripheral blood of acute myeloid leukaemia (AML) patients treated *in vivo* (Fig. 1c, Extended Data Fig. 2 and Extended Data Table 2). Lenalidomide treatment resulted in decreased protein levels of both wild-type isoforms of CK1α as well as two somatic CK1α mutations recently identified in del(5q) MDS patients[15] (Extended Data Fig. 3). Lenalidomide decreased CK1α protein levels without altering *CSNK1A1* mRNA expression (Fig. 1d and Extended Data Fig. 2c), consistent with a post-translational mechanism of regulation. The lenalidomide-dependent decrease in CK1α protein level was abrogated by treatment with the proteasome inhibitor MG132 and the NEDD8-activating enzyme inhibitor

[1]Brigham and Women's Hospital, Division of Hematology, Boston, Massachusetts 02115, USA. [2]University Hospital of Ulm, Department of Internal Medicine III, 89081 Ulm, Germany. [3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [4]Celgene Corporation, San Diego, California 92121, USA. [5]Roswell Park Cancer Institute, Buffalo, New York 14263, USA.
*These authors contributed equally to this work.
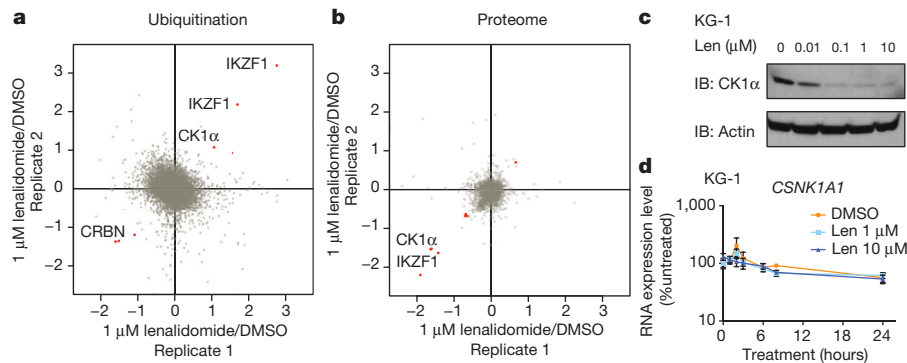
Figure 1 | Lenalidomide-induced changes in ubiquitination and protein levels. a, $\log_2$ ratios for individual K-ε-GG sites of lenalidomide- (1 μM) versus DMSO-treated KG-1 cells for biological replicates 1 and 2. Each point represents a unique K-ε-GG site. Significantly regulated sites ($P < 0.05$) are red. **b**, $\log_2$ ratios of protein abundance for lenalidomide- (1 μM) versus DMSO-treated KG-1 cells for biological replicates 1 and 2. Each point represents a unique protein group. Significantly regulated proteins ($P < 0.05$) are red. **c**, Effects of lenalidomide (Len) on endogenous CK1α protein levels in KG-1 cells after 24-h treatment. Data are representative of 5 independent experiments ($n = 5$). IB, immunoblot. **d**, *CSNK1A1* mRNA levels in KG-1 cells following lenalidomide treatment. Data are mean ± s.d., $n = 3$ biological replicates.

MLN4924, which interferes with the activity of cullin-RING E3 ubiquitin ligases, implicating proteasome- and cullin-dependent degradation of CK1α (Fig. 2a). Homozygous genetic inactivation of CRBN by CRISPR-Cas9 genome editing eliminated lenalidomide-dependent degradation of CK1α, demonstrating CRBN-dependent degradation of CK1α (Fig. 2b and Extended Data Fig. 2d).

We next examined whether CK1α binds CRBN and is ubiquitinated by the CRL4[CRBN] E3 ubiquitin ligase. We observed co-immunoprecipitation of CK1α with endogenous and Flag-tagged CRBN only in the presence of lenalidomide (Fig. 2c and Extended Data Fig. 2e). Lenalidomide treatment increased the ubiquitination of endogenous CK1α in KG-1 cells (Fig. 2d) and in the presence of CRBN *in vitro* (Fig. 2e), confirming that CK1α is a direct target of CRL4[CRBN]. Using a chimaeric protein of CK1α and CK1ε, which shares significant homology with CK1α but is not responsive to lenalidomide, we found that the amino-terminal half (amino acids 1–177) of CK1α is essential for lenalidomide-induced degradation (Extended Data Fig. 3d, e). Sequence alignment with the previously delineated lenalidomide-responsive degron in IKZF1/IKZF3 did not reveal any evident homology, suggesting that CK1α and IKZF1/IKZF3 may interact with the CRBN–lenalidomide complex in distinct manners.

Effect of *CSNK1A1* expression level

We next explored the biological effects of CK1α depletion. CK1α is a serine/threonine kinase with multiple cellular activities. Most notably, CK1α inhibits p53 through MDM2 and MDMX and negatively regulates Wnt signalling as a component of the β-catenin destruction complex[17–21]. In a haematopoietic-specific conditional knockout mouse model, homozygous inactivation of *Csnk1a1* induces apoptosis via p53 activation, while heterozygous loss of *Csnk1a1* causes β-catenin accumulation and stem cell expansion[15]. Similarly, cells haploinsufficient for *Csnk1a1* preferentially undergo apoptosis in response to the casein kinase 1 inhibitor D4476[16]. Since del(5q) cells express about 50% of normal levels of *CSNK1A1* (ref. 14), these results led us to hypothesize that del(5q) cells would be more sensitive to the effects of lenalidomide-induced degradation of CK1α compared to normal cells with two copies of the gene.

To evaluate whether decreased *CSNK1A1* expression sensitizes cells to lenalidomide, we transduced primary human CD34[+] haematopoietic stem and progenitor cells with green fluorescent protein (GFP)-tagged lentiviral vectors expressing *CSNK1A1* or control short hairpin RNAs. Cells expressing *CSNK1A1* shRNAs were depleted in the absence of treatment, demonstrating that knockdown of
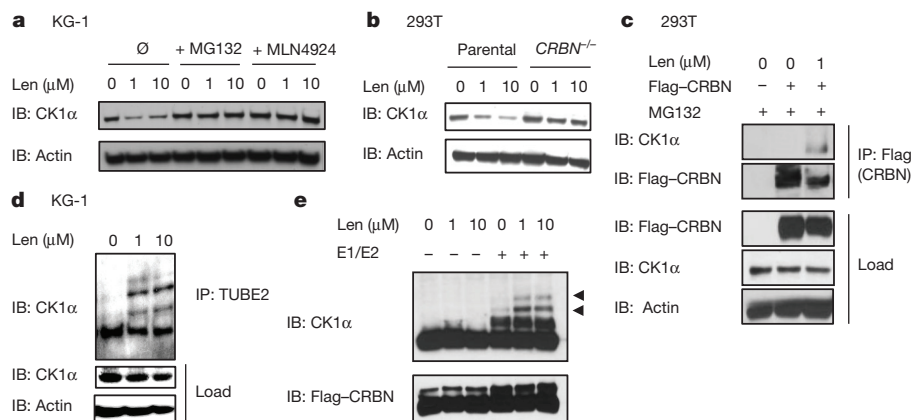
Figure 2 | Lenalidomide induces the ubiquitination of CK1α by CRL4[CRBN]. a, CK1α protein levels in KG-1 cells treated with DMSO or lenalidomide alone or in the presence of 10 μM MG132 or 1 μM MLN4924 for 6 h. **b**, CK1α protein levels in CRBN knockout 293T cells treated with lenalidomide. **c**, Immunoprecipitation of Flag–CRBN in 293T cells treated with DMSO or 1 μM lenalidomide in the presence of 1 μM MG132. IP, immunoprecipitation. **d**, Ubiquitination of endogenous CK1α in KG-1 cells treated with DMSO or lenalidomide analysed by TUBE2 pull-down of ubiquitinated proteins followed by staining with a CK1α-specific antibody. Higher molecular weight bands represent ubiquitinated CK1α. **e**, Ubiquitination of CK1α by CRBN *in vitro* using lysine-free ubiquitin. Arrowheads indicate ubiquitinated CK1α. Results are representative of two (**a, b, d**, $n = 2$) or three independent experiments (**c, e**, $n = 3$). Uncropped blots shown in Supplementary Fig. 1.
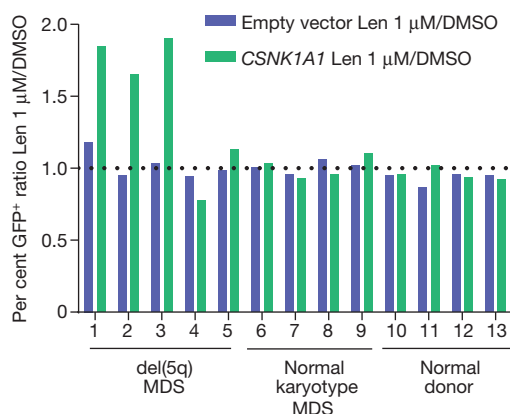
Figure 3 | Ectopic *CSNK1A1* overexpression reduces lenalidomide sensitivity in primary MDS del(5q) cells. CD34$^+$ cells derived from patient or control bone marrow were transduced with a lentiviral vector overexpressing *CSNK1A1* and *GFP* or an empty control vector and treated with DMSO or 1 μM lenalidomide. Results are reported as a ratio of the percentage of GFP$^+$ cells in the lenalidomide condition to the percentage of GFP$^+$ cells in the DMSO condition after 5 days of treatment. A ratio greater than 1 for the *CSNK1A1* vector but not for the empty vector indicates that *CSNK1A1* expression reduces lenalidomide sensitivity. Further information about the patients is given in Extended Data Fig. 5d.

CSNK1A1 inhibits the growth or survival of haematopoietic cells (Extended Data Fig. 4). Treatment with lenalidomide enhanced the depletion of cells expressing *CSNK1A1* shRNAs but had no effect on cells expressing control shRNAs, demonstrating that reduced *CSNK1A1* levels sensitize haematopoietic cells to lenalidomide.

We next evaluated whether overexpression of *CSNK1A1* could reduce the lenalidomide sensitivity of del(5q) MDS cells. We obtained bone marrow samples from MDS patients with heterozygous deletions of chromosome 5q, including heterozygous deletion of *CSNK1A1*, before treatment with lenalidomide. We isolated CD34$^+$ cells from these samples and transduced them with a GFP-tagged lentivirus expressing *CSNK1A1* complementary DNA or empty vector (Fig. 3 and Extended Data Fig. 5). Overexpression of *CSNK1A1* reduced the lenalidomide sensitivity of CD34$^+$ cells from three out of

five del(5q) MDS patients, which compares well with the clinically-observed cytogenetic response rate of about 50%[5–7]. In contrast, over-expression of *CSNK1A1* had no effect on CD34$^+$ cells from patients with normal karyotype MDS or normal donors. Although lenalidomide also induced the degradation of IKZF1 in myeloid cells (Fig. 1a, b), overexpression of *IKZF1* had a similar effect on del(5q) MDS, normal karyotype MDS, and normal donor CD34$^+$ cells, suggesting that degradation of IKZF1 does not explain the therapeutic window of lenalidomide in del(5q) MDS (Extended Data Fig. 5). These findings demonstrate that increased expression of *CSNK1A1* specifically rescues del(5q) cells from lenalidomide treatment.

Species–specific effects of lenalidomide

We next sought to use a conditional knockout mouse model to determine whether haploinsufficiency for *Csnk1a1* sensitizes cells to lenalidomide treatment[15]. In initial experiments, we found that lenalidomide did not decrease CK1α protein levels in mouse Ba/F3 cells, primary murine leukaemia cells, or mice treated *in vivo* (Fig. 4a, b and Extended Data Fig. 6a–d), suggesting that mouse cells are intrinsically resistant to IMiD compounds. Consistent with these findings, mice do not develop the limb malformations observed in human embryos exposed to thalidomide[22] and murine multiple myeloma cells do not respond to lenalidomide[23]. Since CRBN is the direct protein target of lenalidomide, we examined whether expression of human CRBN could confer lenalidomide sensitivity to mouse cells. Overexpression of human, but not mouse, CRBN in mouse Ba/F3 cells resulted in a lenalidomide-dependent decrease of CK1α protein levels, implying that amino acid differences between mouse Crbn and human CBRN are responsible for the species-specific response to lenalidomide (Fig. 4a, b and Extended Data Fig. 6c, d).

To identify the amino acids responsible for this difference, we tested human/mouse CRBN chimaeric proteins for their ability to confer lenalidomide-induced CK1α degradation in mouse Ba/F3 cells (Fig. 4a). Lenalidomide sensitivity was determined by the carboxy-terminal half of CRBN, which contains only 5 non-conserved amino acids between human and mouse. When these non-conserved positions in human CRBN were substituted with the corresponding amino acid in mouse Crbn, only one substitution, V387I (human CRBN isoform 2), disrupted the lenalidomide-responsiveness of
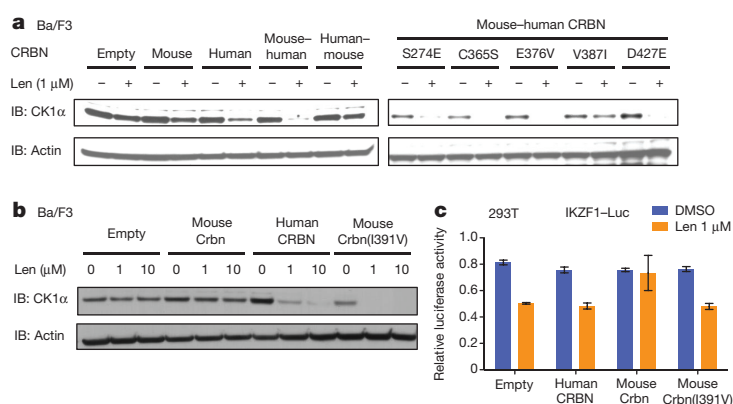
Figure 4 | Amino acid changes in CRBN explain species-specific lenalidomide effects. a, Effect of the expression of human CRBN, mouse Crbn, chimaeras of human and mouse CRBN (mouse–human and human–mouse, breakpoint at residue 221 (human)/225 (mouse)) and variants of the mouse–human chimaera where single amino-acids in the C terminus were mutated to their corresponding mouse residue on lenalidomide-dependent CK1α degradation in mouse Ba/F3 cells. **b**, Expression of mouse *Crbn*^{I391V} restores lenalidomide-dependent CK1α degradation in mouse Ba/F3 cells. See also Extended Data Fig. 6d. **c**, Effect of CRBN, Crbn and Crbn(I391V) on lenalidomide sensitivity of an IKZF1-luciferase fusion protein expressed in human 293T cells. Data are mean ± s.e.m. (*n* = 3 biological replicates).
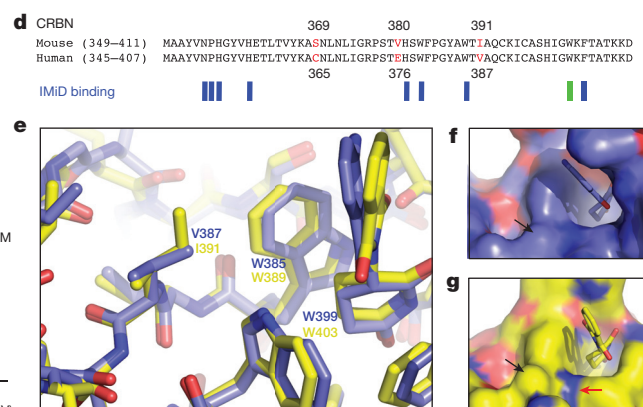
d, Alignment of human and mouse CRBN IMiD binding region. Non-conserved amino acids are red. Amino acids involved in IMiD binding[24,25] are indicated by blue bars. Mouse W403 is indicated with a green bar. **e**, Superposition of the IMiD binding domains of human CRBN (blue, PDB accession 4TZ4) and mouse Crbn (yellow, PDB accession 4TZC). Residues are labelled according to human isoform 2 (blue numbers) and mouse isoform 2 (yellow numbers). **f**, The V387 residue is indicated on the surface of human CRBN with a black arrow. **g**, The corresponding mouse residue, I391, is indicated on the surface of mouse Crbn with a black arrow. Mouse W403 is indicated by a red arrow. Results are representative of 3 (**a**, **b**, **c**) independent experiments. Uncropped blots are shown in Supplementary Fig. 1.
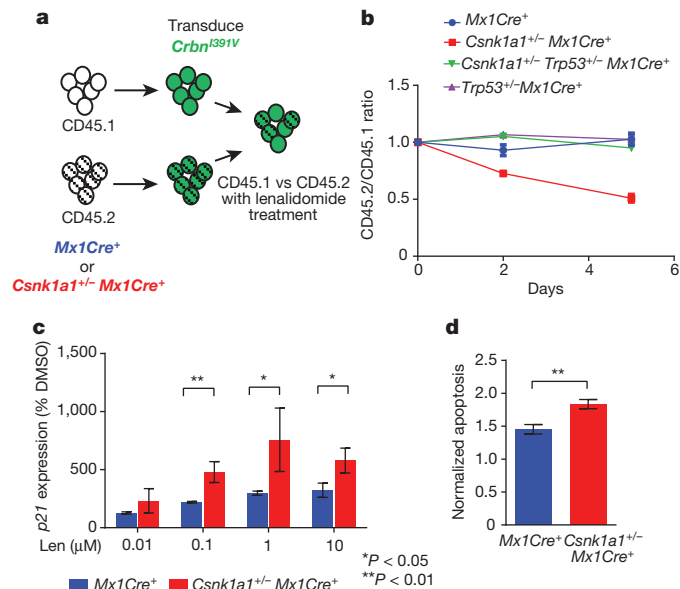
Figure 5 | Effects of lenalidomide treatment on *Csnk1a1*[+/−] mouse haematopoietic cells. **a**, *Csnk1a1*[+/−]*Mx1Cre*[+] or *Mx1Cre*[+] c-Kit[+] haematopoietic stem and progenitor cells (CD45.2) and competitor cells (CD45.1) were transduced with *Crbn*[I391V], mixed in equal ratios, and treated with lenalidomide or DMSO. The relative percentage of CD45.1[+] and CD45.2[+] cells was followed by flow cytometry over 5 days. **b**, Effects of 0.1 µM lenalidomide on the chimaerism of *Csnk1a1*[+/−] *Mx1Cre*[+], *Csnk1a1*[+/−] *Trp53*[+/−] *Mx1Cre*[+], *Trp53*[+/−] *Mx1Cre*[+], or *Mx1Cre*[+] cells (CD45.2) transduced with *Crbn*[I391V] in comparison to CD45.1 competitor cells. Data are shown as mean ± s.e.m., *n* = 3 biological replicates. **c**, Quantitative RT–PCR analysis of *p21* expression in *Csnk1a1*[+/−] or control cells transduced with *Crbn*[I391V] and treated with DMSO or lenalidomide. Data are normalized to DMSO and shown as mean ± s.d., *n* = 3 biological replicates. **d**, Ratio of CD45.2[+] cells and CD45.1[+] cells in late apoptosis (Annexin V[+] DAPI[+]) after transduction with *Crbn*[I391V] and four day treatment with 0.1 µM lenalidomide. CD45.2[+] cells are either *Csnk1a1*[+/−] *Mx1Cre*[+] or *Mx1Cre*[+]. Data are normalized to DMSO treatment. Data are mean ± s.e.m., *n* = 4 biological replicates. Results for **b**, **c**, and **d** are representative of three independent experiments with *hCRBN* or *Crbn*[I391V]. *P* values are from an unpaired two-sided *t*-test.

human CRBN (Fig. 4a). Substitution of the isoleucine at this position in mouse Crbn for the human valine (Crbn(I391V)) was sufficient to confer lenalidomide-induced CK1α degradation in mouse cells (Fig. 4b and Extended Data Fig. 6c, d). Similar effects of these two point mutants were observed on lenalidomide-induced degradation of IKZF1 and IKZF3 in human 293T cells (Fig. 4c and Extended Data Fig. 6e, f), suggesting that a single amino acid change in CRBN determines lenalidomide-responsiveness for multiple substrates.

We modelled the effects of this mouse-human amino acid substitution based on recently published crystal structures of the CRBN–DDB1–IMiD drug complex[24,25]. V387 of human CRBN (equivalent to I391 of mouse Crbn) is located in the IMiD drug binding region of CRBN, but does not directly interact with lenalidomide (Fig. 4d). To investigate how the substitution of isoleucine for valine in mouse Crbn confers lenalidomide-responsiveness, we superimposed the structures for the mouse and human IMiD-binding regions bound to lenalidomide as solved in Chamberlain *et al.* (2014)[25]. No backbone changes are present at the site of the valine-isoleucine species differences (Fig. 4e), but the isoleucine residue is well-defined in the electron density with the long arm of the side chain oriented towards the indole NH moiety of W403 in the mouse structure (Extended Data Fig. 7). The increase in steric bulk of the isoleucine side chain, relative to valine, results in a bulge in the solvent accessible surface of the mouse protein adjacent to both W403 and lenalidomide (Fig. 4f, g). It has been proposed that IMiD binding produces a hotspot for substrate interactions by placement of the hydrophobic phthalimide or isoindolinone ring in an environment of potential hydrogen bond donors and acceptors from the surface of CRBN[25]. In this case, the larger side chain of the isoleucine residue found in rodents may sterically clash with substrate proteins such as IKZF1 and CK1α, blocking access to key hydrogen bonds from CRBN, such as from the indole NH from tryptophan 403 (mouse numbering). Steric clashes and occlusion of key bonds with substrate proteins thereby provides a potential explanation of why IMiD compounds bind mouse Crbn[1] but do not promote degradation of IKZF1 and CK1α.

Having determined the mechanism of lenalidomide resistance in mouse cells, we expressed the *Crbn*[I391V] cDNA in haematopoietic cells from *Csnk1a1* conditional knockout mice to determine the effects of *Csnk1a1* haploinsufficiency on lenalidomide sensitivity. We isolated
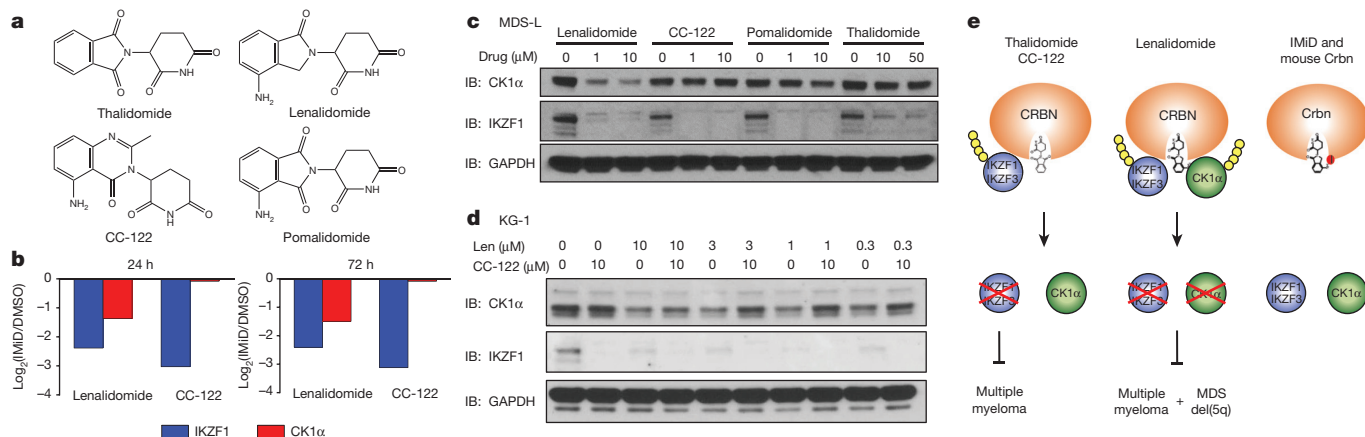
Figure 6 | Substrate specificity of thalidomide analogues. **a**, Structures of thalidomide, lenalidomide, pomalidomide and CC-122. **b**, Protein levels of IKZF1 and CK1α assessed by tandem mass tag quantitative proteomics in MDS-L cells treated with DMSO, 10 µM lenalidomide, or 1 µM CC-122 for 24 h (left panel) or 72 h (right panel). *n* = 3 (drug treatment) or *n* = 4 (DMSO). **c**, Western blot analysis of CK1α and IKZF1 protein levels in MDS-L cells treated with DMSO or different concentrations of thalidomide, lenalidomide, pomalidomide, or CC-122. Results are representative of two independent experiments (*n* = 2). **d**, KG-1 cells were treated with DMSO or lenalidomide in the absence or presence of different concentrations of CC-122. Results are representative of five independent experiments in various cell lines. **e**, Schematic presentation of the interaction of different thalidomide analogues with CRBN, substrates, and therapeutic indications.

CD45.2$^+$ c-Kit$^+$ haematopoietic stem and progenitor cells from *Csnk1a1*$^{+/-}$*Mx1Cre*$^+$ and *Mx1Cre*$^+$ control littermates treated with poly (I:C) to induce gene excision in haematopoietic cells. We transduced these cells with a retroviral vector expressing *Crbn*^{I391V}, and cultured them in competition with similarly transduced isogenic CD45.1 c-Kit$^+$ cells in the presence or absence of lenalidomide (Fig. 5a). Lenalidomide had no effect on control cells, but *Csnk1a1*$^{+/-}$*Mx1Cre*$^+$ cells were significantly depleted in the presence of lenalidomide (Fig. 5b). The enhanced sensitivity of *Csnk1a1*$^{+/-}$*Mx1Cre*$^+$ cells to lenalidomide was associated with induction of the p53 target gene *p21* (Fig. 5c) and increased levels of apoptosis (Fig. 5d), and was rescued by heterozygous deletion of *Trp53* (Fig. 5b), demonstrating a critical down-stream role for the p53 pathway. These results are consistent with the clinical observation that *TP53* mutations confer lenalidomide resistance in MDS with del(5q)[26].

Differential substrate specificity

Thalidomide, lenalidomide, and pomalidomide target IKZF1 and IKZF3 for ubiquitination and degradation and are active in multiple myeloma, but only lenalidomide has been shown to be clinically effective in del(5q) MDS[2,5–7]. We therefore asked whether different thalidomide analogues induce degradation of the same substrates. We used tandem mass tag (TMT) quantitative proteomics[27] in the MDS-L cell line to compare the activities of lenalidomide and CC-122, a novel CRBN-binding agent that shares the glutarimide ring and has recently entered clinical trials (Fig. 6a). As expected, treatment with lenalidomide significantly decreased protein levels of both IKZF1 and CK1α. In striking contrast, treatment with 1 μM CC-122 caused an even greater decrease in IKZF1 ($P = 2.77 \times 10^{-10}$, 72 h) than 10 μM lenalidomide ($P = 2.10 \times 10^{-8}$, 72 h), but had no effect on CK1α protein levels ($P > 0.05$) (Fig. 6b and Extended Data Fig. 8a, b).

We confirmed the TMT mass spectrometry findings by western blot for IKZF1 and CK1α in MDS-L and KG-1 cells (Fig. 6c and Extended Data Fig. 8c, d). While all compounds induced degradation of IKZF1, thalidomide and CC-122 did not affect CK1α protein levels, even at high concentrations, and pomalidomide had only weak effects on CK1α protein levels. Although CC-122 has a greater potency than lenalidomide for degradation of IKZF1 and IKZF3, it was ineffective in decreasing CK1α protein levels compared to lenalidomide, suggesting that subtle chemical modifications can affect substrate preference (Fig. 6a–c). Furthermore, treatment with excess CC-122 abrogated the lenalidomide-induced degradation of CK1α, demonstrating that lenalidomide and CC-122 compete for the same glutarimide binding site on CRBN (Fig. 6d). Consistent with the role of CK1α as a negative regulator of Wnt signalling, we observed increased levels of β-catenin after treatment with lenalidomide but not CC-122 or thalidomide (Extended Data Fig. 8e–g). These experiments demonstrate that despite structural similarity, the substrate specificities of thalidomide analogues differ. Notably, only lenalidomide has a strong effect on CK1α, suggesting that it may indeed be most appropriate modulator of CRL4^{CRBN} for the treatment of del(5q) MDS (Fig. 6e).

Intriguingly, lenalidomide, but not thalidomide or pomalidomide, has been reported to induce the formation of two β-strands composed of CRBN residues 346–363[25]. Although conformational differences are difficult to interpret in the absence of a substrate-bound structure, the formation of these β-strands is expected to make significant changes in the surface of CRBN near the IMiD binding site[25] and thus it may contribute to the differential recruitment of IKZF1 and CK1α. The interaction of specific thalidomide analogues with particular substrates may therefore be governed by unique structural determinants, revealing the biological and clinical potential for members of this class of drugs to induce degradation of distinct sets of proteins.

Discussion

We demonstrate that lenalidomide targets CK1α for degradation, and that heterozygous deletion of *CSNK1A1* in del(5q) MDS provides a therapeutic window for selective targeting of the malignant cells by lenalidomide. The concept that genes within heterozygous deletions could cause vulnerabilities in cancer cells was first proposed 20 years ago[28] and has been more recently demonstrated as CYCLOPS genes (cancer vulnerabilities unveiled by genomic loss) in cell lines[29]. Our data demonstrate that in del(5q) MDS, lenalidomide-induced degradation of CK1α below haploinsufficient levels induces p53 activity and growth inhibition, as CK1α is a negative regulator of p53. Deletion of contiguous genes on chromosome 5q, such as *RPS14*, may further sensitize del(5q) cells to p53 activation[9,30,31]. This mechanism of activity is consistent with the acquisition of *TP53* mutations in del(5q) MDS patients who develop resistance to lenalidomide. Degradation of CK1α may also contribute to other clinical effects of lenalidomide such as activity in the activated B-cell (ABC) subtype of diffuse large B-cell lymphoma[32] and lenalidomide-induced myelosuppression. Further investigation is required to determine the complete biological effects from degradation of each substrate.

Lenalidomide, like thalidomide and pomalidomide, binds CRBN and induces degradation of specific substrates. We found that a single amino acid difference between mouse and human CRBN renders mouse cells insensitive to IMiD compounds. This discovery enabled us to demonstrate, using a genetically engineered mouse model, that *Csnk1a1* haploinsufficiency sensitizes cells to lenalidomide. Non-conserved amino acid changes in CRBN may also explain why thalidomide does not cause teratogenicity in mice and was approved for use in pregnant women, leading to the birth of more than 10,000 newborns with limb malformations and other disabilities.

Thalidomide, lenalidomide, and pomalidomide all induce CRL4^{CRBN}-mediated degradation of IKZF1 and IKZF3, but the subtle differences in chemical structure between these molecules cause dramatic changes in potency. We now find that thalidomide and a novel compound, CC-122, induce the degradation of IKZF1 but not CK1α. CC-122 may have a greater therapeutic window for the treatment of B cell malignancies and other diseases that depend on IKZF1 and IKZF3, but would not be predicted to have activity in del(5q) MDS.

CC-122, like thalidomide and its analogues, has a glutarimide ring that anchors the molecule in CRBN, and structural variation in the remainder of the molecule is thought to determine substrate specificity[24,25]. These findings provide evidence that thalidomide-related molecules have distinct biological activities, mediated by degradation of distinct sets of substrates, and that these compounds will be the first in a larger class of drugs with therapeutic utility through the targeting of specific proteins for degradation.

1. Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).
2. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
3. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
4. Gandhi, A. K. *et al.* Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4^{CRBN}. *Br. J. Haematol.* **164**, 811–821 (2014).
5. List, A. *et al.* Lenalidomide in the myelodysplastic syndrome with chromosome 5q deletion. *N. Engl. J. Med.* **355**, 1456–1465 (2006).
6. List, A. *et al.* Efficacy of lenalidomide in myelodysplastic syndromes. *N. Engl. J. Med.* **352**, 549–557 (2005).
7. Fenaux, P. *et al.* A randomized phase 3 study of lenalidomide versus placebo in RBC transfusion-dependent patients with low-/intermediate-1-risk myelodysplastic syndromes with del5q. *Blood* **118**, 3765–3776 (2011).
8. Pellagatti, A. *et al.* Lenalidomide inhibits the malignant clone and up-regulates the *SPARC* gene mapping to the commonly deleted region in 5q- syndrome patients. *Proc. Natl Acad. Sci. USA* **104**, 11406–11411 (2007).

9. Wei, S. et al. A critical role for phosphatase haplodeficiency in the selective suppression of deletion 5q MDS by lenalidomide. Proc. Natl Acad. Sci. USA **106,** 12974–12979 (2009).

10. Boultwood, J. et al. Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. Blood **99,** 4638–4641 (2002).

11. Graubert, T. A. et al. Integrated genomic analysis implicates haploinsufficiency of multiple chromosome 5q31.2 genes in de novo myelodysplastic syndromes pathogenesis. PLoS ONE **4,** e4583 (2008).

12. Ong, S. E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics **1,** 376–386 (2002).

13. Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. Nature Protocols **8,** 1950–1960 (2013).

14. Boultwood, J. et al. Gene expression profiling of CD34$^+$ cells in patients with the 5q− syndrome. Br. J. Haematol. **139,** 578–589 (2007).

15. Schneider, R. K. et al. Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. Cancer Cell **26,** 509–520 (2014).

16. Järås, M. et al. Csnk1a1 inhibition has p53-dependent therapeutic efficacy in acute myeloid leukemia. J. Exp. Med. **211,** 605–612 (2014).

17. Huart, A. S., MacLaine, N. J., Meek, D. W. & Hupp, T. R. CK1α plays a central role in mediating MDM2 control of p53 and E2F–1 protein stability. J. Biol. Chem. **284,** 32384–32394 (2009).

18. Wu, S., Chen, L., Becker, A., Schonbrunn, E. & Chen, J. Casein kinase 1α regulates an MDMX intramolecular interaction to stimulate p53 binding. Mol. Cell. Biol. **32,** 4821–4832 (2012).

19. Elyada, E. et al. CKIα ablation highlights a critical role for p53 in invasiveness control. Nature **470,** 409–413 (2011).

20. Knippschild, U. et al. The CK1 family: contribution to cellular stress response and its role in carcinogenesis. Front. Oncol. **4,** 96 (2014).

21. Venerando, A., Ruzzene, M. & Pinna, L. A. Casein kinase: the triple meaning of a misnomer. Biochem. J. **460,** 141–156 (2014).

22. Fratta, I. D., Sigg, E. B. & Maiorana, K. Teratogenic effects of thalidomide in rabbits, rats, hamsters, and mice. Toxicol. Appl. Pharmacol. **7,** 268–286 (1965).

23. Chesi, M. et al. Drug response in a genetically engineered mouse model of multiple myeloma is predictive of clinical efficacy. Blood **120,** 376–385 (2012).

24. Fischer, E. S. et al. Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. Nature **512,** 49–53 (2014).

25. Chamberlain, P. P. et al. Structure of the human Cereblon–DDB1–lenalidomide complex reveals basis for responsiveness to thalidomide analogs. Nature Struct. Mol. Biol **21,** 803–809 (2014).

26. Jädersten, M. et al. TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. J. Clin. Oncol. **29,** 1971–1979 (2011).

27. Rauniyar, N. & Yates, J. R. III. Isobaric labeling-based relative quantification in shotgun proteomics. J. Proteome Res. **13,** 5293–5309 (2014).

28. Frei, E. III. Gene deletion: a new target for cancer chemotherapy. Lancet **342,** 662–664 (1993).

29. Nijhawan, D. et al. Cancer vulnerabilities unveiled by genomic loss. Cell **150,** 842–854 (2012).

30. Ebert, B. L. et al. Identification of RPS14 as a 5q⁻ syndrome gene by RNA interference screen. Nature **451,** 335–339 (2008).

31. Narla, A. & Ebert, B. L. Ribosomopathies: human disorders of ribosome dysfunction. Blood **115,** 3196–3205 (2010).

32. Yang, Y. et al. Exploiting synthetic lethality for the therapy of ABC diffuse large B cell lymphoma. Cancer Cell **21,** 723–737 (2012).

Supplementary Information is available in the online version of the paper.

Author Contributions J.K., E.C.F., and B.L.E. initiated the project. J.K., E.C.F., P.W.H., K.J.M., S.N.H, M.McC., A.K.G., M.J., and R.K.S. designed and performed cell experiments and protein analysis. E.C.F. designed and performed mouse competition and patient sample experiments. N.D.U., D.R.M., and T.S. performed KG-1 proteomics and analysis. E.G. and M.W. provided patient samples. P.P.C. performed the structural analysis. H.W.M. synthesized CC-122. K.J.M. provided MDS-L proteomics. B.E.C., R.C., L.B., S.A.C., and B.L.E. supervised the work. J.K., E.C.F., and B.L.E. wrote the manuscript. All authors assisted in editing the manuscript.

METHODS

The experiments were not randomized, and no statistical methods were used to predetermine sample size.

Reagents. Lenalidomide (Toronto Research Chemicals, Selleck Chemicals, and Celgene), Thalidomide (Millipore and Celgene), Pomalidomide (Selleck Chemicals and Celgene), MG-132 (Selleck Chemicals), CC-122 (Celgene), PR619 (Lifesensors), MLN4924 (Active Biochem), and Leptomycin B (Santa Cruz) were dissolved in DMSO at 10 to 100 mM and stored at $-20\,°C$ for up to 6 months. For cell culture experiments drugs were diluted at least by 1:1,000 so that the final DMSO concentration was 0.1% or lower.

Cell lines. KG-1, Ba/F3, K562, MM1S, Jurkat, HEL, and 293T cells were obtained from American Type Culture Collection (ATCC) and their identity was not further authenticated. MDS-L cells were provided by Kaoru Tohyama, Kawasaki Medical School (Japan). Cells were cultured in RPMI 1640 (Mediatech) or DMEM (Mediatech) supplemented with 10–20% heat-inactivated fetal bovine serum (FBS)(Omega Scientific) and 1% penicillin, streptomycin, and L-glutamine (Mediatech). Cells were grown at $37\,°C$ in a humidified incubator under 5% CO_2. Ba/F3 cells were cultured in the presence of 10 ng ml^{-1} mouse IL-3 (Miltenyi) and MDS-L cells were cultured with 10 ng ml^{-1} human GM-CSF. 293T cells were transfected using TransIT-LT1 (Mirus Bio) according to the manufacturer's protocol. Cell lines were intermittently tested for mycoplasma.

Cell culture and treatment for K-ε-GG and proteome profiling. KG-1 cells were cultured for 2 weeks (~6 cell doublings) in RPMI depleted of L-arginine and L-lysine (Caisson Labs Inc.) and supplemented with 10% dialysed FBS (Sigma) and L-arginine (Arg0) and L-lysine (Lys0) (light), $^{13}C_6^{14}N_4$-L-arginine (Arg6) and 4,4,5,5-D$_4$-L-lysine (Lys4) (medium) or $^{13}C_6^{15}N_4$-L-arginine (Arg10) and $^{13}C_6^{15}N_2$-L-lysine (Lys8) (heavy) to generate light-, medium- and heavy-labelled cells. Media was exchanged every 3rd day. On day 14 cells were treated with 1 μM lenalidomide, 10 μM lenalidomide or DMSO for 4 h for ubiquitination profiling and 24 h for protein level assessment. Experiments were performed in two biological replicates with flipped SILAC labelling: replicate 1: DMSO/light, lenalidomide 1 μM/medium; lenalidomide 10 μM/heavy; replicate 2: lenalidomide 1 μM/light; lenalidomide 10 μM/medium; DMSO/heavy.

SILAC based K-ε-GG and proteome profiling of KG-1 cells. Cell lysis and trypsin digestion, basic pH reversed phase fractionation, K-ε-GG enrichment, and LC-MS/MS analysis for KG-1 cells were performed as recently described[2] with minor changes. Cell pellets used for K-ε-GG profiling were lysed in 8 M urea, 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 2 μg ml^{-1} aprotinin (Sigma-Aldrich), 10 μg ml^{-1} leupeptin (Roche Applied Science), 1 mM phenyl-methylsulfonyl fluoride (PMSF), 50 μM PR-619, and 1 mM chloroacetamide at $4\,°C$. Pellets used for proteome profiling were lysed in 8 M urea, 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 2 μg ml^{-1} aprotinin (Sigma-Aldrich), 10 μg ml^{-1} leupeptin (Roche Applied Science), 1 mM phenylmethylsulfonyl fluoride (PMSF), at $4\,°C$. For this work, 10 mg of protein was input per SILAC state for the ubiquitin workflow. For proteome profiling, 1.5 mg of protein was input per SILAC state. Proteins were reduced with 5 mM dithiothreitol for 45 min at room temperature and subsequently carbamidomethylated with 10 mM iodoacetamide for 30 min at RT in the dark. Samples were diluted to 2 M urea with 50 mM Tris-HCl, pH 7.5, and digested with sequencing grade trypsin (Promega) at $25\,°C$ overnight using an enzyme-to-substrate ratio of 1:50. Digested samples were acidified to 1% formic acid (Sigma-Aldrich). Tryptic peptides were centrifuged for 5 min at 3,000g to remove precipitate. Peptides were desalted exactly as previously described[2].

Samples were fractionated by basic pH reversed phase (bRP) fractionation using an Agilent 1100 Series HPLC and Zorbax 300 Å Extend-C18 columns as previously described[13,33]. A 9.4 mm × 250 mm column (Agilent, 5 μm bead size) was used to fractionate samples intended for K-ε-GG enrichment, whereas a 4.6 mm × 250 mm column (Agilent, 3.5 μm bead size) was used to fractionate samples intended for proteome analysis[33]. For K-ε-GG samples, approximately 15 mg of peptide sample was resuspended in 1.8 ml of basic RP solvent A (2% MeCN, 5 mM ammonium formate, pH 10), separated into 2 HPLC vials and injected with Solvent A at flow rate of 3 ml min^{-1}. A 64-min method was used for fractionation exactly as previously described[2,13]. A total of 96 2 ml fractions were collected every 0.66 min at a flow rate of 3 ml min^{-1}. After separation, bRP fractions were pooled in a serpentine, noncontiguous manner to generate 8 final fractions (final fraction 1 = 1, 9, 17, 25, 33, 41, 49, 57, 65; final fraction 2 = 2, 10, 18, 26, 34, 42, 50, 58, 66; …). Since 10 mg of protein per SILAC state was used for K-ε-GG samples, and the maximum loading capacity on the 9.4 mm × 250 mm column bRP column is 15 mg, two rounds of fractionation were completed per replicate sample.

Proteome samples were brought up in 0.9 ml of basic RP solvent A and injected with solvent A at flow rate of 1 ml min^{-1}. Peptides were separated using the bRP method previously described for the 4.6 mm × 250 mm column[33]. Approximately 5% of each proteome fraction was taken and pooled to generate 24 final fractions (final fraction 1 = 1, 25, 49; final fraction 2 = 2, 26, 50; …) plus a fraction "A" that contains early eluting peptides. All bRP pooled fractions were dried using a SpeedVac concentrator.

K-ε-GG enrichment was completed using the anti-K-ε-GG antibody obtained from the PTMScan ubiquitin remnant motif (K-ε-GG) kit (Cell Signaling Technology) as previously described[2,13]. Briefly, bRP fractions we reconstituted in 1.5 ml of immunoaffinity purification buffer and each fraction was incubated with 31 μg of cross-linked anti-K-ε-GG antibody for 1 h at $4\,°C$ with end-over-end rotation. Following incubation, samples were spun down and the supernatant was removed. Antibody-bound beads were washed 4× with 1.5 ml of ice cold PBS. Peptides were eluted with 2× 50 μl of 0.15% trifluoroacetic acid (TFA). Eluted peptides were desalted using C18 StageTips.

K-ε-GG and proteome fractions were reconstituted in 9 μl and 20 μl of 3% MeCN/1% FA, respectively, and analysed using a Q Exactive mass spectrometer (Thermo Fisher Scientific) coupled on-line to a Proxeon Easy-nLC 1000 system. For analysis of each fraction, 4 μl and 1 μl of K-ε-GG and global proteome samples was injected, respectively. Samples were injected onto a microcapillary column (360 μm outer diameter × 75 μm internal diameter) packed with 24 cm of ReproSil-Pul C18-AQ 1.9 μm beads (Dr. Maisch GmbH) that was heated to $50\,°C$ and equipped with an integrated electrospray emitter tip (10 μm). For online LC separation, solvent A was 0.1% FA/3% MeCN and solvent B was 90% MeCN/0.1% FA. Peptides were eluted into the mass spectrometer using the liquid chromatography-mass spectrometry (LC-MS) method previously described[2]. The Q Exactive instrument was operated in the data-dependent mode acquiring 12 HCD MS/MS scans ($R = 17,500$) after each MS1 scan ($R = 70,000$) using an MS1 ion target of $3 × 10^6$ ions and an MS2 target of $5 × 10^4$ ions. The maximum ion time for the MS/MS scans was set to 120 ms, the collision energy was set to 25, the dynamic exclusion time was set to 20 s, and the peptide match setting was set to on.

The MaxQuant software version 1.3.0.5 was used to analyse MS data. Data was searched against the human Uniprot database as well as a database provided by MaxQuant containing common laboratory contaminants. The search parameters were as follows: enzyme specificity was set to trypsin, maximum number of mixed cleavages set to 2, precursor mass tolerance was set to 20 ppm for the first search, and set to 6 ppm for the main search. Oxidized methionine and N-terminal protein acetylation were searched as variable modifications and carbamido-methylation of cysteine was searched as a fixed modification. Data files from K-ε-GG enriched samples were also searched with Gly-Gly addition to lysine as a variable modification. The minimum peptide length was set to 6, and false discovery rate for peptide, protein, and site identification was set to 1%. Reverse and contaminant hits were removed from data sets. Normalized ratios were used for quantification. For proteome data, proteins identified by 2 or more razor/unique peptides and quantified by 2 or more ratio counts in both biological replicates were considered for the final data set. For the K-ε-GG data, K-ε-GG sites were considered if they were quantified in both biological replicates.

For data analysis, normalized SILAC ratios for the 2 biological replicates were filtered to retain only those deemed reproducible. Reproducibility was based on replicates being confined within the 95% limits of agreement of a Bland–Altman plot[34]. In the Bland–Altman plot, differences of the replicates are plotted against the average values and the limits of agreement correspond to the prediction confidence interval for a regression line with unit slope. Reproducible replicates were then subjected to a moderated t-test to assess statistical significance[35]. This statistic is similar to the ordinary t-statistic, with the exception that the standard errors are calculated using an empirical Bayes method using information across all proteins, thereby making inference about each individual protein more robust. The nominal P values arising from the moderated t-statistic are corrected for multiple testing by controlling the false discovery rate (FDR), as proposed by Benjamini and Hochberg[36]. Proteins with an FDR adjusted P value of less than 0.05 were deemed to be reproducibly regulated. Figures containing scatter plots of SILAC data show all points regardless of the reproducibility measure. Statistical significance was assessed using only reproducible data points.

The original mass spectra may be downloaded from MassIVE (http://massive. ucsd.edu) using the identifier: MSV000079014. The data are accessible at ftp://massive.ucsd.edu/MSV000079014.

Plasmids and virus constructs. The following cDNAs were cloned in the pRSF91 retrovirus backbone (gift of C. Baum, Hanover Medical School) or pEF1a-IRES-GFP lentiviral backbone: CSNK1A1 Isoform 2 (ccsbBroadEN_06055), CSNK1A1 Isoform 1 (gift from W. G. Kaelin), CSNK1E (ccsbBroadEN_00379), mouse Crbn Isoform 2 (Thermo Scientific), and human CRBN Isoform 2 (ccsbBroadEn_08244). Human IKZF1 isoform 1 was synthesized using gBlocks (IDT) with internal BstXI and BsrGI sites removed using synonymous substitutions. For certain

experiments GFP was replaced by GFP-T2A-PAC (*Puromycin* N-acetyl-tranfer-ase gene) to allow for drug selection with puromycin of positively transduced cells. Chimeric cDNAs and point mutations were cloned with overlapping PCR primers. Point mutations in mouse-human CRBN chimaeric proteins are annotated according to their position in human CRBN isoform 2. Lentivirus was concentrated by ultracentrifugation for transduction of primary cells.

Lentiviral vectors (TRC005 backbone) expressing shRNAs targeting luciferase (TRCN0000072254: ATGTTTACTACACTCGGATAT) and CSNK1A1 (#1: TRCN0000342505, CATCTATTTGGCGATCAACAT; #2: TRCN0000342507, GCAGAATTTGCGATGTACTTA) were obtained from The RNAi Consortium (TRC) of the Broad Institute. For certain experiments, the PAC gene was replaced by GFP.

The luciferase reporter plasmid pCMV-IRES-RenillaLUC-IRES-Gateway-FireflyLUC[3] was a gift from W. G. Kaelin (Dana-Farber Cancer Institute). Cloning of cDNAs was performed using Gateway LR reaction (Invitrogen).

CRISPR mediated genetic deletion was performed with the pSgRNA-CAS9-T2A- PAC plasmid[37] using CRBN exon 1-specific guide RNAs (CRBN targeting sequence: #1 TCCTGCTGATCTCCTTCGC, #2 AACCACCTGCCGCTCCT GCC). 1×10^5 293T cells were transfected in a 12-well with 500 ng of each plasmid using TransLTI (Mirius). After 24 h, transfected cells were selected with $2 \mu g \, ml^{-1}$ puromycin for 4 days. Then 293T cells were diluted to single cell and plated in 96-well. Colonies were tested by western blot and Sanger sequencing of the endogenous CRBN exon 1 locus for inactivating biallelic out-of-frame mutations.

Western blot and antibodies. Protein lysates were run on Tris-HCl, 1 mm Criterion Precast gels (Bio-Rad) or NuPAGE Bis-Tris gels (Novex) gels at a constant voltage. Proteins were transferred onto Immobilon-P transfer membranes (Millipore) at a constant amperage. Before staining with primary antibodies, blots were blocked in 5% non-fat dry milk (Santa Cruz) or 5% BSA in TBS-T 0.1% for 30 min.

For protein detection primary antibodies detecting CK1α (C-19, Santa Cruz or Abcam ab108296), β-catenin (Cell Signaling #9587 and #8480), haemagglutinin (HA; horseradish peroxidase (HRP)-conjugate, Miltenyi, GG8-1F3.3), Flag (M2, HRP-conjugate Sigma Aldrich), ubiquitin conjugates (FK2, HRP-conjugate Enzo Life Sciences), actin (HRP-conjugate, Abcam), β-tubulin (Cell Signaling #2146) and GAPDH (Santa Cruz sc-47724) were used. Secondary antibodies were HRP conjugated bovine anti-goat (Jackson ImmunoResearch) and HRP-conjugated donkey anti-rabbit (GE Healthcare). SuperSignal (Thermo Scientific) chemiluminescent substrate was used for detection. For re-probing, blots were stripped in Restore Western Blot Stripping Buffer (Thermo Scientific), activated in methanol, and re-blocked.

Flow cytometry. Flow cytometry was performed on a FACS Canto II (BD Bioscience) using the PE and FITC channels for the detection of dTomato and GFP, respectively. DAPI staining was performed to exclude dead cells. A High-Throughput Sampler (BD) was used for some experiments.

Quantitative RT–PCR. Gene expression was measured by reverse transcription quantitative PCR (RQ-PCR). For RNA isolation and reverse transcription a cDNA Synthesis Kit for MultiMacs (Miltenyi) was used according to the manufacturer's protocol. The following primer-probe sets from Life Technologies were used with TaqMan Gene Expression Master Mix (Life Technologies): human *GAPDH* (402869), human *CSNK1A1* (Hs00793391_m1), human *IKZF1* (Hs00958474_m1), mouse *GAPDH* (Mm99999915_g1), mouse *p21* (Mm04205640_g1). Analysis was performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems) in a 384-well plate. Relative expression levels were calculated using the $\Delta\Delta C_T$ method.

Immunoblot analysis of patient samples. Frozen viable patient samples from the CC-5013-AML-001 trial (http://clinicaltrials.gov identifier NCT01358734) were thawed at 37 °C, washed with PBS, and cell pellets were frozen at −80 °C. Cells were lysed in RIPA buffer containing HALT Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific), quantified with a BCA Protein Assay Kit and 3–5 μg of protein was run on a Bis-Tris gel (Novex). Membranes were stained with anti-CK1α and anti-GAPDH antibodies and detected with chemiluminescence. Informed consent was obtained from all subjects in this trial, including consent to use the collected materials to study the mechanism of lenalidomide and its effects on specific proteins. Samples were collected according to IRB-approved protocols at the 30 sites at which this study was conducted. Due to limited availability of patient samples, this experiment could be performed only once.

Immunoprecipitation of CRBN. For immunoprecipitation of Flag–CRBN, 3×10^6 293T cells were plated in a 10 cm dish and transfected with 10 μg pRSF91-Flag-hCRBN or empty vector. Cells were treated with DMSO or 1 μM lenalidomide in the presence of 10 μM MG132 for 3 h. Cells were lysed in Pierce IP Lysis Buffer and lysates were cleared by centrifugation. Flag–CRBN

was immunoprecipitated overnight using anti-Flag M2 Affinity Gel (Sigma-Aldrich) in the presence of 10 μM MG132 and DMSO or 1 μM lenalidomide. The beads were washed 3 times with IP lysis buffer (Pierce) and protein was eluted from the affinity gel with 250 μg ml^{-1} Flag peptide (Sigma) after incubation for 30 min at 4 °C. Protein lysates were then analysed as described above.

For immunoprecipitation of endogenous CRBN 5×10^6 293T cells were treated with DMSO or 10 μM lenalidomide and 10 μM MG132 for 4 h. Protein lysates were incubated overnight at 4 °C with 1 μg of a polyclonal mouse anti-CRBN antibody (abcam) in the presence of lenalidomide or DMSO and MG132, Protein G Sepharose beads were added for one hour. The beads were washed once with IP lysis buffer (Pierce) and protein was eluted from the beads by incubation with LDS loading buffer (Life Technologies) at 70 °C for 10 min.

***In vivo* ubiquitination.** For assessment of endogenous ubiquitination of CK1α 2×10^7 KG-1 cells were treated with DMSO, 1 or 10 μM lenalidomide for 4 h and then lysed in IP lysis buffer containing 10 mM NEM and 10 μM MG132. Ubiquitinated proteins were pulled down by Ubiquilin 1 Tandem UBA (TUBE2) Agarose (Boston Biochem) for 4 h at 4 °C and washed 3× with IP lysis buffer. Protein was eluted by incubation with Laemmli buffer (Biorad) at 95 °C for 5 min, separated by SDS–PAGE, transferred to PVDF membrane and probed with anti-CK1α.

***In vitro* ubiquitination.** 293T cells were transfected with either HA-CK1α or Flag–CRBN expressing vectors. After 48 h, cells were lysed in Pierce IP lysis buffer (Thermo Scientific) and immunoprecipitated overnight with Flag–Sepharose beads (Anti-Flag M2 Affinity Gel, Sigma) or HA-Sepharose beads (EZView Red anti-HA affinity gel, Sigma). The beads were washed 3× in IP lysis buffer and 2× in E3 Ligase Reaction buffer (Boston Biochem) and eluted with 250 μg ml^{-1} Flag peptide (Sigma) or 100 μg ml^{-1} HA peptide for 30 min at 4 °C. The eluates were mixed in a 1:1 ratio and added to a ubiquitination reaction mixture containing 200 nM E1 (UBE1), 2 μM UbcH5a, 1 μM UbcH5c, 1 μg l^{-1} K$_0$ ubiquitin, 1 μM ubiquitin aldehyde, 1× Mg-ATP, 1× E3 Ligase Reaction Buffer (all Boston Biochem), 10 μM MG132, 100 nM MG101 and 1 μM lenalidomide, 10 μM lenalidomide, or DMSO (1:1,000) as appropriate in a total volume of 25 μl.

Negative controls did not include E1 and E2 enzymes. After a 90 min incubation at 30 °C, the reaction was denatured by adding 5× SDS containing loading buffer (Boston Biochem), boiled at 95 °C for 5 min, separated by SDS–PAGE and transferred to a PVDF membrane in order to detect HA-CK1α and its ubiquitinated forms with a CK1α-specific antibody. The membrane was then stripped and re-probed with anti-Flag antibody.

Immunofluorescence. 50,000 293T cells were grown on Lab-Tek 8 well chamber slides (Nunc) for 24 h and then treated with DMSO or 10 μM lenalidomide for various durations. At the conclusion of treatment, the media was decanted and the wells were washed 1× with PBS. Cells were fixed in 4% formaldehyde in PBS for 15 min, washed 3× 5 min in PBS and blocked for 1 h at room temperature in PBS with 0.3% Tween-20 and 5% BSA. Primary antibody was anti-CK1α (C-19, Santa Cruz), which was diluted 1:100 in PBS with 0.3% Tween-20 and 1% BSA (antibody dilution buffer) and incubated for 2 h at room temperature. After 3 × 5 min washes in PBS, Alexa Fluor 488 donkey anti-Goat (Life Technologies) was added at 1:200 in antibody dilution buffer and incubated for 1 h at room temperature. After 3× 5 min washes in PBS slides were coverslipped with Vectashield mounting media with DAPI (Vector Laboratories). Slides were analysed by fluorescence microscopy at 100× using a Nikon Eclipse 90i and NIS Elements. Channels were merged using ImageJ.

Purification, culture, and lentiviral infection of human CD34$^+$ cells for shRNA experiments. Research cord blood units were obtained from The New York Blood Center according to an Institutional Review Board-approved protocol. Cord blood CD34$^+$ haematopoietic cells were isolated from Ficoll purified PBMCs with an Indirect CD34 MicroBead kit (Miltenyi) and an Auto MACS Pro (Miltenyi) according to the manufacturer's protocol. Cells were cultured in serum free media (SFEM, StemSpan) containing 50 ng ml^{-1} recombinant human TPO (Miltenyi), 40 ng ml^{-1} human FLT3 ligand (Miltenyi), 25 ng ml^{-1} recombinant human SCF (Miltenyi), and 10 ng ml^{-1} IL-3 (Miltenyi). For shRNA experiments, CD34$^+$ cells were transduced with a VSV-G pseudotyped TRC pLKO.005 lentiviral vector expressing GFP instead of the puromycin resistance gene. Infection was performed after 24 h in culture in a 96-well using spinfection in the presence of 2 μg ml^{-1} polybrene (hexadimethrine bromide, Sigma). 48 h after transduction the number of transduced cells was analysed by flow cytometry and was used as baseline. Then cells were cultured in 1 μM lenalidomide or DMSO and the relative number of infected cells was assessed by flow cytometry for 3 weeks.

Purification, culture, and lentiviral infection of patient samples. Viably frozen bone marrow mononuclear cells were obtained from healthy donors or patients with del(5q) MDS according to IRB approved protocols at the University of Pennsylvania and Roswell Park Cancer Institute. Informed consent was obtained

from all subjects. Samples were thawed and CD34$^+$ haematopoietic cells were isolated 20–24 h later using an Indirect CD34 MicroBead kit (Miltenyi) and an Auto MACS Pro (Miltenyi). Cells were grown in serum free media (SFEM, StemSpan) supplemented with 25 ng ml^{-1} SCF, 40 ng ml^{-1} FLT3 ligand, 50 ng ml^{-1} thrombopoietin, 40 μg ml^{-1} lipids, 100 U ml^{-1} Pen/Strep and 2 mM glutamine. 6–8 h after CD34$^+$ isolation, cells were transduced with concentrated VSV-G pseudotyped pEF1a-GFP-IRES-hCSNK1A1, pEF1a-GFP-IRES-hIKZF1 or empty vector control virus via spinfection in the presence of 4 μg ml^{-1} polybrene (Sigma, diluted to 2 μg ml^{-1} after spinfection). After 3 days, the initial percentage of transduced cells was determined by flow cytometry and remaining cells were split to treatment with either DMSO or 1 μM lenalidomide. The relative abundance of transduced cells in each condition was assessed after 5 days by flow cytometry. Control cord-blood CD34$^+$ cells were isolated as above. Adult bone marrow CD34$^+$ cells were purchased as single-donor lots from AllCells (Alameda, CA). The number of replicates for each patient sample, vector, and treatment was limited by the number of cells available and was as follows: one for samples 1–5, 8, 9 and 12; two for samples 6, 7 and 10; three for sample 11; four for sample 13. Samples were combined from three experiments.

For qPCR validation of *CSNK1A1* or *IKZF1* mRNA expression, cord blood CD34$^+$ cells were transduced with lentivirus expressing GFP and the cDNA of interest or empty vector. After 3 days, transduced GFP$^+$ cells were FACS sorted and RNA extraction and qPCR was performed as above.

***TP53* sequencing of patient samples.** Genomic DNA was extracted from the CD34$^-$ fraction of the patient bone marrow samples using a DNA Blood Mini Kit (Qiagen). PCR and sequencing was performed as described in the International Agency for Research on Cancer's Direct Sequencing Protocol (http://p53.iarc.fr/download/tp53_directsequencing_iarc.pdf). Mutations were identified using Mutation Surveyor (Softgenetics). Benign polymorphisms were identified using the International Agency for Research on Cancer's Portal (http://p53.iarc.fr/TP53GeneVariations.aspx).

Expressing different CRBN proteins in Ba/F3 cells. Variants of human and mouse CRBN were cloned into a modified pRSF91 backbone to generate pRSF91-CRBN-IRES-GFP-T2A-PAC retroviral constructs. 200,000 Ba/F3 cells were infected with ecotropic retrovirus in the presence of 2 μg ml^{-1} polybrene. After 24 h, 1 μg ml^{-1} puromycin (Gibco) was added and cells were selected for 3–4 days. Cells were confirmed to be >90% GFP$^+$ by flow cytometry and 1,000,000 cells were plated per 6-well and treated with DMSO or lenalidomide for 24 h. Protein lysates were harvested and immunoblotted for CK1α as described above.

IKZF1 and IKZF3 luciferase reporter assay. 10,000 293T cells were transfected with 50 ng of pCMV-IRES-RenillaLUC-IRES-IKZF1/IKZF3-FireflyLUC reporter plasmid together with 100 ng of a vector expressing CRBN, Crbn, different chimaeric or mutant CRBN forms or empty control. After 48 h, cells were treated with DMSO and lenalidomide for 4 h. Firefly and *Renilla* luciferase activity was measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol.

Mouse experiments. Mouse experiments were performed according to an IACUC approved protocol at Children's Hospital Boston. Generation and characterization of the conditional *Csnk1a1* knockout mouse has been described previously[15]. *Csnk1a1*$^{flox/+}$ mice were crossed with *Mx1Cre*$^+$ mice on a C57BL/6NTac background to obtain *Csnk1a1*$^{flox/+}$ *Mx1Cre*$^+$ mice. *Csnk1a1*$^{flox/+}$ *Mx1Cre*$^+$ or control *Csnk1a1*$^{+/+}$ *Mx1Cre*$^+$ mice were treated with 3 doses of 200 μg poly(I:C) (Invivogen HMW) at 8-10 weeks of age and gene excision was confirmed where applicable. At least 2 weeks following poly(I:C) treatment, the long bones and spines were harvested and crushed and red blood cells were lysed. Bone marrow from age-matched mice of the same genotype was pooled to create sex-balanced groups. C-Kit$^+$ cells were isolated with a CD117 MicroBead Kit (Miltenyi) and an AutoMacs Pro and grown in SFEM (StemSpan) supplemented with antibiotics and 50 ng ml^{-1} mTPO (Peprotech) and 50 ng ml^{-1} mSCF (Peprotech) for 24 h. Ecotropic pseudotyped retrovirus was spun onto Retronectin (Clontech) coated 6-well plates and cells were added in 1 ml of media with 2 μg ml^{-1} polybrene. An additional 1 ml of media was added after 24 h. After 48 h, GFP$^+$ cells were isolated by FACS sorting (BD FACS Aria II) and CD45.1 and CD45.2 cells were mixed. Cells were treated with various doses of lenalidomide and the percent CD45.1 and CD45.2 cells expressing the fluorescent marker was followed by flow cytometry over time following cell surface staining. Antibodies for flow cytometry were as follows: CD45.1 APC/Cy7 (A20, BioLegend), CD45.2 PE (104, eBioscience), and CD45.2 FITC (104, eBioscience). The number of mice needed for each experiment was calculated by assuming a yield of 5 million c-Kit$^+$ cells per mouse (long bones and spine), a transduction rate of 40%, and a target of 50,000 GFP$^+$ cells per experimental well, which was found to reduce variability in pilot experiments. Due to the experimental design, the genotypes of the mice could not be blinded or randomized.

For *in vivo* treatment, female C57BL/6 mice were treated with once daily intraperitoneal injection of 10 mg per kg lenalidomide dissolved in DMSO and diluted in 100 μl PBS or with DMSO in 100 μl PBS. Bone marrow was harvested after 14 days and lysed with IP lysis buffer (Pierce). Due to experimental design, the treatment groups could not be blinded.

Quantitative mass spectrometry in MDS-L cells. MDS-L multiplexed quantitative mass spectrometry samples were processed and analysed by the Thermo Fisher Scientific Center for Multiplexed Proteomics at Harvard Medical School. Samples were prepared as previously described[38] with the following modification. All solutions are reported as final concentrations. Lysis buffer (8 M urea, 1% SDS, 50 mM Tris pH 8.5, protease and phosphatase inhibitors from Roche) was added to the cell pellets to achieve a cell lysate with a protein concentration between 2–8 mg ml^{-1}. A micro-BCA assay (Pierce) was used to determine the final protein concentration in the cell lysate. Proteins were reduced and alkylated as previously described. Proteins were precipitated using methanol/chloroform. In brief, four volumes of methanol was added to the cell lysate, followed by one volume of chloroform, and finally three volumes of water. The mixture was vortexed and centrifuged to separate the chloroform phase from the aqueous phase. The precipitated protein was washed with one volume of ice cold methanol. The washed precipitated protein was allowed to air dry. Precipitated protein was resuspended in 4 M urea, 50 mM Tris pH 8.5. Proteins were first digested with LysC (1:50; enzyme:protein) for 12 h at 25 °C. The LysC digestion is diluted down to 1 M urea, 50 mM Tris pH 8.5 and then digested with trypsin (1:100; enzyme:protein) for another 8 h at 25 °C. Peptides were desalted using a C$_{18}$ solid phase extraction cartridges as previously described. Dried peptides were resuspended in 200 mM EPPS, pH 8.0. Peptide quantification was performed using the micro-BCA assay (Pierce). The same amount of peptide from each condition was labelled with tandem mass tag (TMT) reagent (1:3; peptide:TMT label) (Pierce). The 6-plex and 10-plex labelling reactions were performed for 2 h at 25 °C. Modification of tyrosine residue with TMT was reversed by the addition of 5% hydroxyl amine for 15 min at 25 °C. The reaction was quenched with 0.5% TFA and samples were combined at a 1:1:1:1:1:1 ratio for 6-plex experiments or 1:1:1:1:1:1:1:1:1:1 for 10-plex experiments. Combined samples were desalted and offline fractionated into 24 fractions as previously described.

Liquid chromatography-MS3 spectrometry (LC-MS/MS) in MDS-L cells. 12 of the 24 peptide fraction from the basic reverse phase step (every other fraction) were analysed with an LC-MS3 data collection strategy on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) equipped with a Proxeon Easy nLC 1000 for online sample handling and peptide separations. Approximately 5 μg of peptide resuspended in 5% formic acid + 5% acetonitrile was loaded onto a 100-μm inner diameter fused-silica micro capillary with a needle tip pulled to an internal diameter less than 5 μm. The column was packed in-house to a length of 35 cm with a C$_{18}$ reverse phase resin (GP118 resin 1.8 μm, 120 Å, Sepax Technologies). The peptides were separated using a 120 min linear gradient from 3% to 25% buffer B (100% acetonitrile (ACN) + 0.125% formic acid) equilibrated with buffer A (3% ACN + 0.125% formic acid) at a flow rate of 600 nl min^{-1} across the column. The scan sequence for the Fusion Orbitrap began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, 400−1,400 m/z scan range, AGC target 2×10^5, maximum injection time 100 ms, dynamic exclusion of 75 s). 'Top speed' (1 s) was selected for MS2 analysis, which consisted of CID (quadrupole isolation set at 0.5 Da and ion trap analysis, AGC 4×10^3, NCE 35, maximum injection time 150 ms). The top ten precursors from each MS2 scan were selected for MS3 analysis (synchronous precursor selection), in which precursors were fragmented by HCD before Orbitrap analysis (NCE 55, max AGC 5×10^4, maximum injection time 150 ms, isolation window 2.5 Da, resolution 60,000 (10-plex experiments) or 15,000 (6-plex experiments)).

LC-MS3 data analysis for MDS-L cells. A suite of in-house software tools were used to for. RAW file processing and controlling peptide and protein level false discovery rates, assembling proteins from peptides, and protein quantification from peptides as previously described[38]. MS/MS spectra were searched against a Uniprot human database (February 2014) with both the forward and reverse sequences. Database search criteria are as follows: tryptic with two missed cleavages, a precursor mass tolerance of 50 ppm, fragment ion mass tolerance of 1.0 Da, static alkylation of cysteine (57.02146 Da), static TMT labelling of lysine residues and N termini of peptides (229.162932 Da), and variable oxidation of methionine (15.99491 Da). TMT reporter ion intensities were measured using a 0.03 Da window (6-plex) or 0.003 Da window (10-plex) around the theoretical m/z for each reporter ion in the MS3 scan. Peptide spectral matches with poor quality MS3 spectra were excluded from quantitation (<100 summed signal-to-noise across 6 channels and <0.5 precursor isolation specificity for 6-plexes or (<200 summed signal-to-noise across 10 channels and <0.5 precursor isolation specificity for 10-plexes).

A moderated *t*-test was applied across all proteins to assess statistical significance[35]. This statistic is similar to the ordinary *t*-statistic, with the exception that the standard errors are calculated using an empirical Bayes method using information across all proteins, thereby making inference about each individual protein more robust. This test assumes normality and uses an estimated standard deviation intended to handle relatively few replicates per condition. Posterior residual standard deviations are used in place of ordinary standard deviation in the moderated *t*-test applied. This shrinkage of protein-wise sample variances to a pooled estimate provides more stable inference when sample numbers are reduced. The nominal *P* values arising from the moderated *t*-statistic are corrected for multiple testing by controlling the false discovery rate (FDR), as proposed by Benjamini and Hochberg[36]. Proteins with an FDR adjusted *P* value of less than 0.05 were deemed to be reproducibly regulated.

IMiD compound substrate selectivity in KG-1 and MDS-L cells. Cells ($2\text{-}4 \times 10^6$) were plated in 10-cm dishes and incubated overnight (18–24 h). Cells were treated with DMSO, lenalidomide (1–10 μM), CC-122 (1–10 μM), pomalidomide (1–10 μM), or thalidomide (10–100 μM) for 6 h. Drug- treated cells were collected, washed with PBS and cell pellets were lysed in RIPA buffer containing protease and phosphatase inhibitors for 30–45 min followed by

sonication and centrifugation. Protein lysates were quantified using BCA protein assay kit and 10–15 μg of protein was used for western analysis. CC-122 and lenalidomide competition experiment was conducted as above except cells were pre-treated with DMSO or 10 μM CC-122 for 90 min followed by treatment with lenalidomide (0.3–10 μM) or DMSO for 6 h.

The synthesis and characterization of CC-122 is described in the Supplementary Methods section.

33. Mertins, P. *et al.* Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nature Methods* **10,** 634–637 (2013).
34. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1,** 307–310 (1986).
35. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3,** Article 3 (2004).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., B* (1995).
37. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnol.* (2014).
38. Weekes, M. P. *et al.* Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* **157,** 1460–1472 (2014).
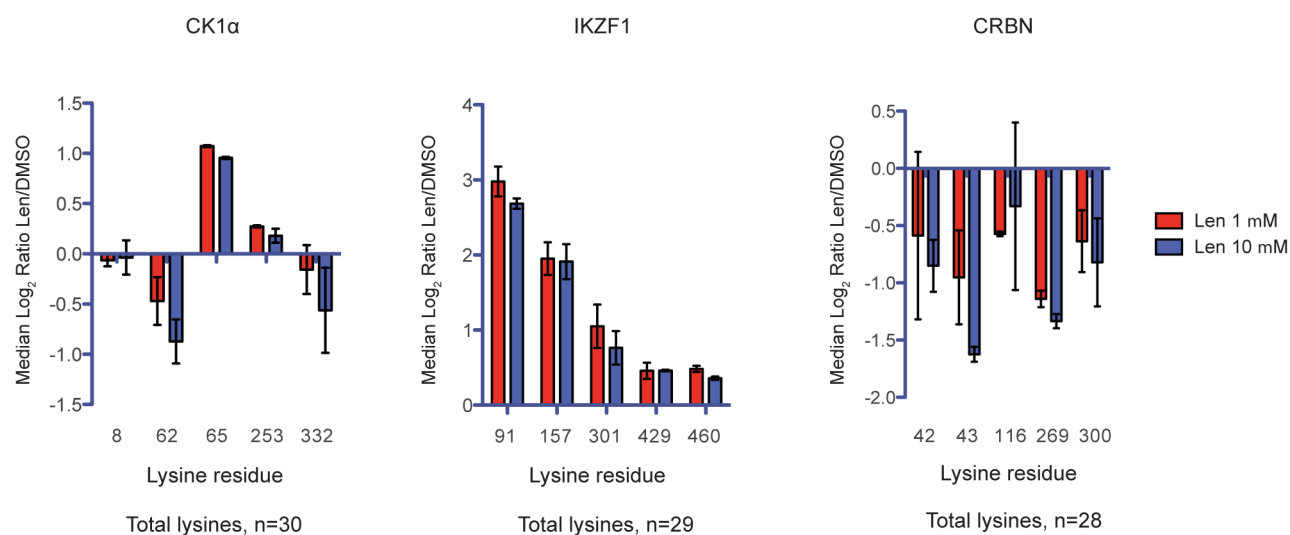
Extended Data Figure 1 | Effect of lenalidomide on specific ubiquitination sites. Median log$_2$ ratios for different lysine residues in CK1α isoform 2, IKZF1 isoform 1, and CRBN isoform 2 for 1 or 10 μM lenalidomide-treated KG-1 cells versus DMSO-treated cells. SILAC experiments were performed in two biological replicates with flipped SILAC labelling. Only lysine residues detected in both replicates are shown. Error bars show range.
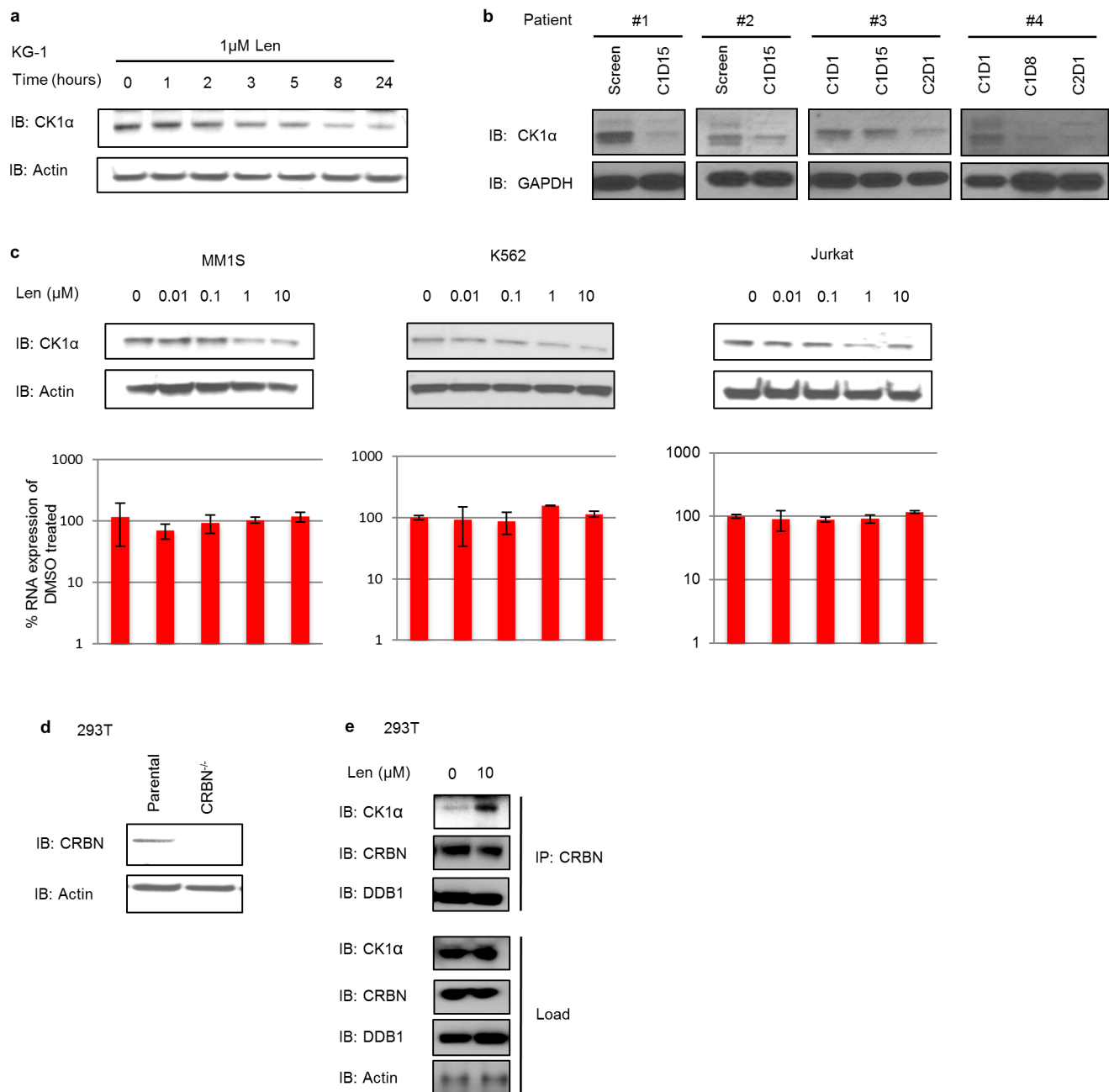
Extended Data Figure 2 | Effect of lenalidomide in human cells. a, Time course of effect of lenalidomide treatment on CK1α protein levels in KG-1 cells. **b,** Immunoblot of CK1α protein levels in the bone marrow (1, 2) and peripheral blood (3, 4) mononuclear cells of AML patients treated with lenalidomide as part of a clinical trial. Pre-treatment samples are taken at the screen or before the first treatment (C1D1). Subsequent time points are cycle 1 day 15 (C1D15), cycle 2 day 1 (C2D1) or cycle 1 day 8 (C1D8) of lenalidomide treatment. Further details about these patients (n = 4) can be found in Extended Data Table 2. **c,** MM1S, K562, and Jurkat cells were treated with different concentrations of lenalidomide for 24 h. CK1α protein levels were detected by western blot and *CSNK1A1* mRNA expression levels were measured by RQ-PCR. Data are mean ± s.d., n = 3 each with three technical replicates. **d,** Immunoblot confirming loss of CRBN expression in 293T cells with the CRBN gene disrupted by CRIPSR-Cas9 genome editing. **e,** Immunoprecipitation with a CRBN-specific antibody in 293T cells treated with DMSO or 10 μM lenalidomide for 5 h in the presence of 10 μM MG132. Results in **a, c, d,** and **e** are each representative of two independent experiments. Uncropped blots are shown in Supplementary Fig. 1.
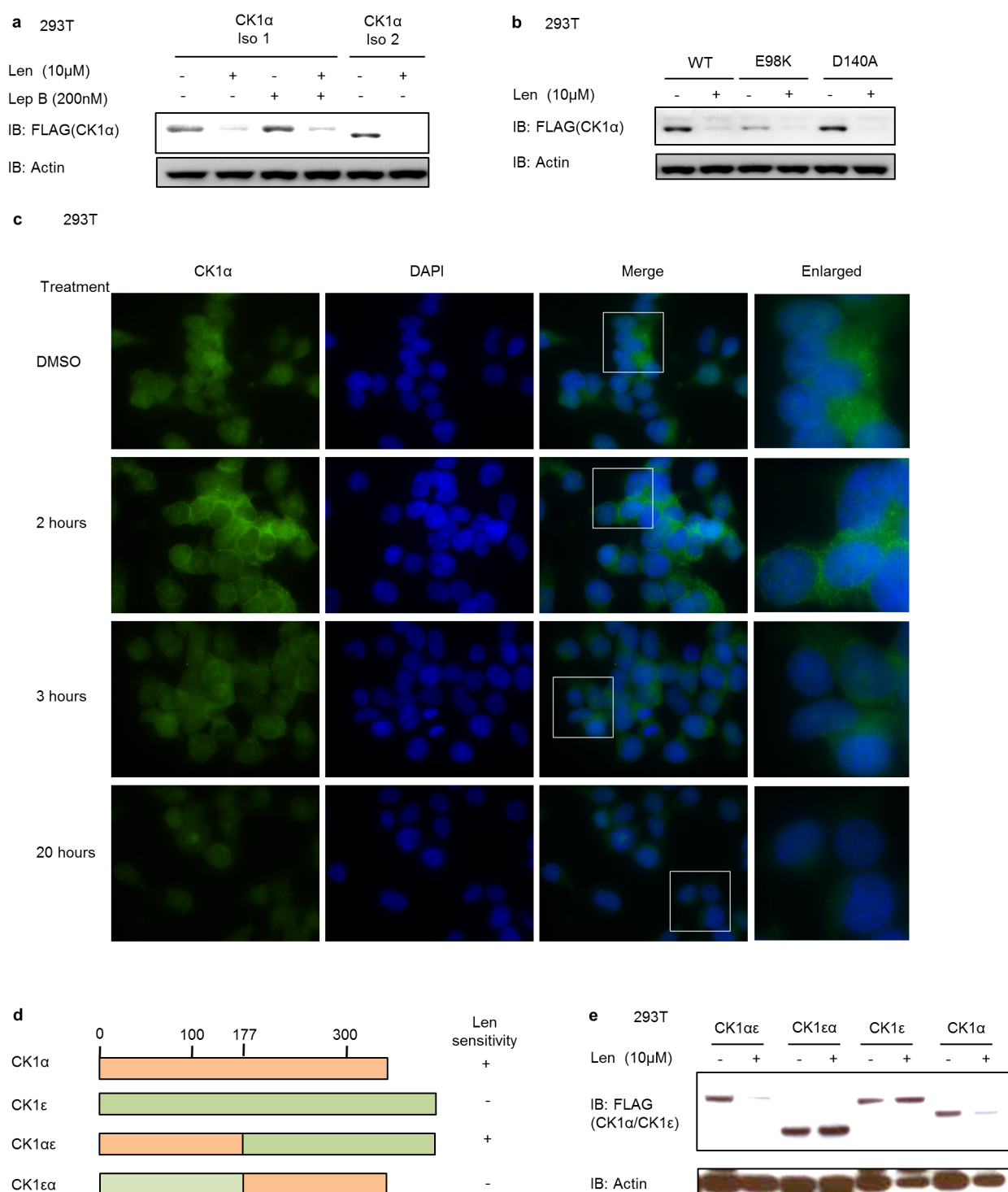
Extended Data Figure 3 | Sequence determinants of CK1α degradation.
a, 293T cells were transfected with plasmids expressing Flag–CK1α isoform 1 or isoform 2 together with a human CRBN-expressing plasmid. Cells were treated with DMSO or 10 μM lenalidomide for 16 h. Cells expressing Flag–CK1α isoform 1, which contains a nuclear localization domain, were incubated in the absence or presence of the nuclear export inhibitor leptomycin B. **b**, 293T cells expressing Flag–CK1α isoform 2 wild-type or two different point mutations identified in patient samples were treated with DMSO or 10 μM lenalidomide for 16 h. **c**, Immunofluorescence for CK1α after treatment with DMSO or 10 μM lenalidomide. Enlarged area is indicated by a box in Merge. FITC channel represents staining for CK1α. No changes in CK1α localization are seen upon lenalidomide treatment. Experiment was performed twice in biological duplicate. In each condition, at least 25 cells were assessed. **d**, Chimaeric proteins of casein kinase 1A1 (CK1α) and casein kinase 1E (CK1ε), which shares significant homology with CK1α but is not responsive to lenalidomide, that were used in **e** to determine the lenalidomide-responsive region in CK1α. **e**, Flag-tagged (chimaeric) proteins from **d** were transfected in 293T cells together with a CRBN-expressing plasmid. Cells were treated with 1 μM lenalidomide for 24 h and protein was detected with a Flag-specific antibody. Data are representative of two (**a**, **c**), three (**b**) or four (**e**) independent experiments. Uncropped blots are shown in Supplementary Fig. 1.
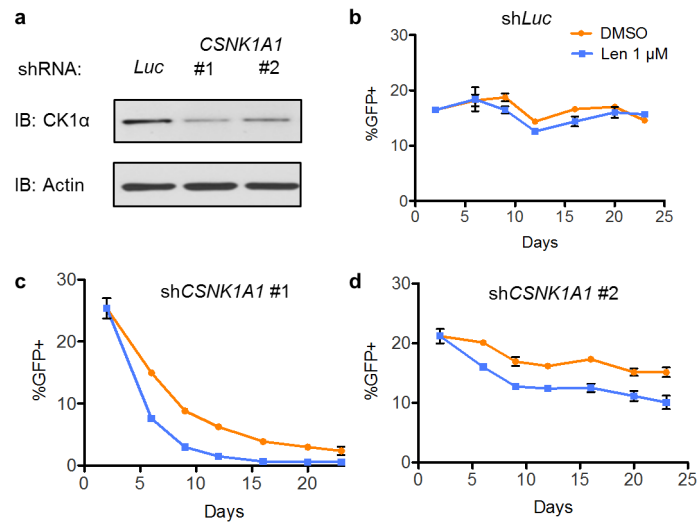
Extended Data Figure 4 | CSNK1A1 knockdown increases lenalidomide sensitivity in haematopoietic cells. a, Knockdown validation by western blot. **b–d**, CD34$^+$ cells were transduced with GFP-labelled lentivirus expressing either control shRNA targeting luciferase (**b**) or shRNA targeting *CSNK1A1* (**c, d**) and treated with DMSO or 1 μM lenalidomide. The percentage of GFP$^+$ cells was assessed by flow cytometry over time. Results are representative of 3 independent experiments each with $n = 3$ biological replicates.

a — **b** bar charts; **c** bar chart

d

Sample #	Karyotype	Category	*TP53* Status (exons sequenced)
1	46,XX,del(5)(q15q33)[14] / 46,XX,del(5)(q15q33),del(12)(p11p13.3)[6] / 47,XX,del(5)(q15q33),+8[1] / 46,XX[3]	del(5q) MDS Pre-Lenalidomide	WT (2-8, 10-11)
2	46,XX,del(5)(q13q33)[17] / 46,XX[3]	del(5q) MDS Pre-Lenalidomide	ND
3	46,XY,del(5)(q15q35) [23] / 46,XY[2]	del(5q) MDS Pre-Lenalidomide	ND
4	46,XX,del(5)(q13q33),t(12;13)(q15;q11) [16]/46,XX,+8[1] / 46,XX [3]	del(5q) MDS Pre-Lenalidomide	WT (2-11)
5	46,XX,del(5)(q13q33) [14] / 46,XX [6]	del(5q) MDS Pre-Lenalidomide	WT (2-11)
6	normal	Normal Karyotype MDS	WT (4-7, 9-10)
7	normal	Normal Karyotype MDS	WT (2-11)
8	normal	Normal Karyotype MDS	WT (2-11)
9	normal	Normal Karyotype MDS	WT (3-4, 6-11)
10	normal	Normal Adult Bone Marrow CD34+	ND
11	normal	Normal Adult Bone Marrow CD34+	ND
12	normal	Normal Adult Bone Marrow CD34+	ND
13	normal	Cord Blood CD34+	ND

ND: *TP53* sequencing not done due to limited sample size or normal donor sample

WT: *TP53* sequence is wild-type (known benign polymorphisms only) in the exons sequenced

Extended Data Figure 5 | Expression of *CSNK1A1* and *IKZF1* in patient samples. a, mRNA expression of *CSNK1A1* in cord blood CD34+ cells infected with lentivirus expressing human *CSNK1A1* or empty vector. CD34+ cells were infected with GFP-tagged lentivirus and GFP+ cells were sorted three days later. Values are mean ± s.d., $n = 4$ biological replicates, each with 3 technical replicates. **b**, mRNA expression of *IKZF1* in cord blood CD34+ cells infected with lentivirus expressing *IKZF1* or empty vector as in **a**. Values are mean ± s.d., $n = 3$ biological replicates, each with 3 technical replicates. **c**, CD34+ cells derived from patient or control bone marrow were transduced with a lentivirus expressing human *IKZF1* (h*IKZF1*) and *GFP* or an empty control vector and treated with DMSO or 1 μM lenalidomide. The percentage of GFP+ cells was assessed by flow cytometry after five days for each vector-drug combination. Results are reported as a ratio of the percentage of GFP+ cells in the lenalidomide condition to the percentage of GFP+ cells in the DMSO condition. Results are combined from three experiments. **d**, Characteristics of patient samples used for *CSNK1A1* and *IKZF1* expression experiments. Results of *TP53* sequencing, including exons with adequate coverage, is given in the rightmost column. All samples sequenced had wild-type *TP53*. ND, not done due to limited patient material. WT, *TP53* exon sequence has only known benign polymorphisms.
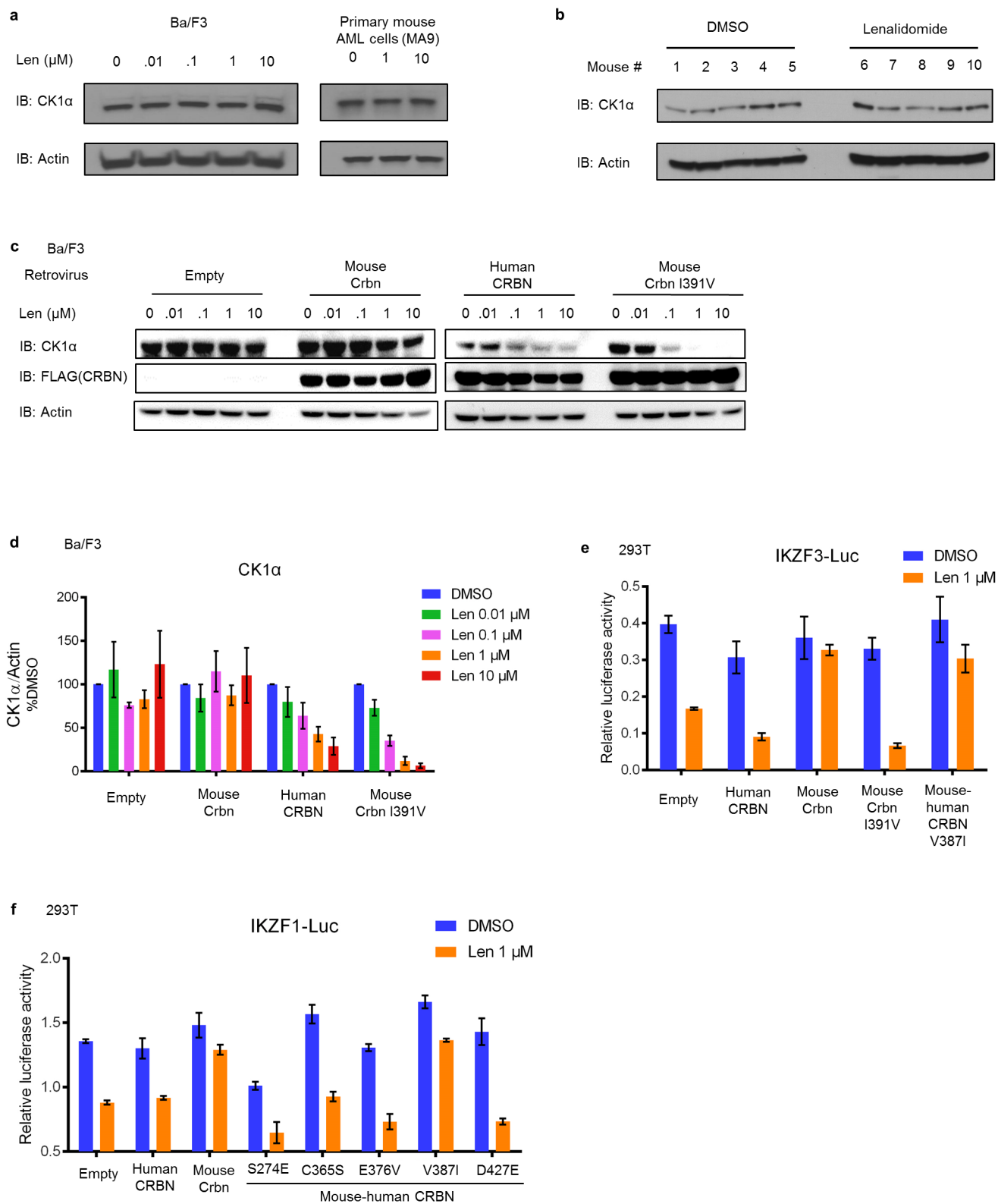
Extended Data Figure 6 | Effect of lenalidomide on mouse cells. a, CK1α protein levels are unaffected in mouse Ba/F3 cells and primary mouse AML cells (MA9) treated with a range of lenalidomide doses. Data are representative of two independent experiments ($n = 2$). **b**, CK1α expression in bone marrow cells of mice treated with DMSO ($n = 5$) or lenalidomide ($n = 5$). **c**, CK1α protein levels in Ba/F3 cells transduced with empty vector, mouse Crbn, human CRBN or Crbn(I391V) and treated with lenalidomide. **d**, Quantification of CK1α protein levels in Ba/F3 cells using ImageJ. Graphs show the fraction of normalized CK1α protein levels as compared to control (DMSO) treated cells of the respective line. Bars represent mean ± s.e.m. from three independent experiments as in **c**. **e**, **f**, Effect of lenalidomide on an IKZF3–luciferase (**e**) and IKZF1–luciferase fusion protein (**f**) in 293T cells expressing human, mouse or different chimaeras or mutations of CRBN. Data are shown as mean ± s.e.m. ($n = 3$, biological replicates) and are representative of three (**f**) or five (**e**) independent experiments. Uncropped blots are shown in Supplementary Fig. 1.
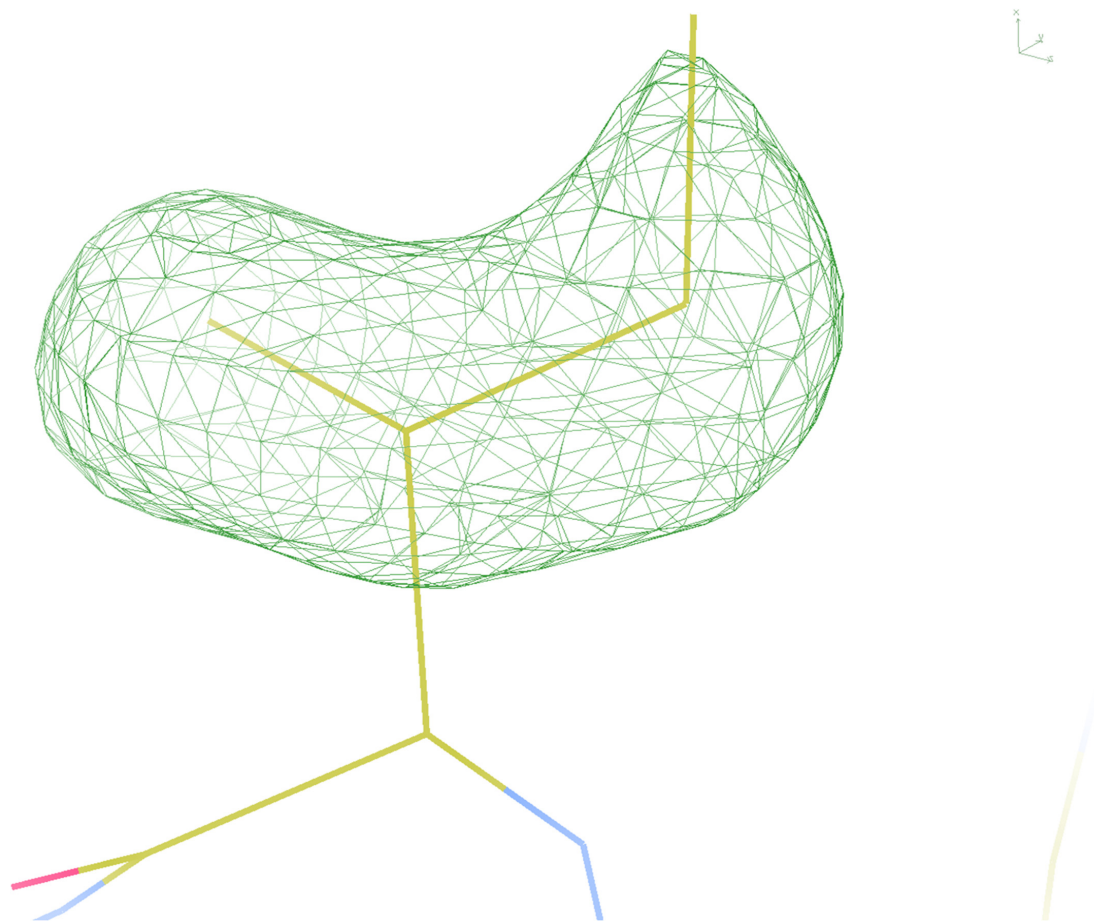
Extended Data Figure 7 | Difference electron density map of mouse residue I391 calculated in the absence of a side chain showing the favoured orientation of the residue. The density is contoured at 3.8σ following a single round of Refmac5 refinement.

a Log₂ fold change

	24 hr		72 hr	
Protein	Len	CC-122	Len	CC-122
CK1α	-1.37	-0.08	-1.49	-0.08
IKZF1	-2.38	-3.03	-2.41	-3.12

b Adjusted P value

	24 hr		72 hr	
Protein	Len	CC-122	Len	CC1-22
CK1α	9.84E-06	0.16	1.49E-07	0.08
IKZF1	4.16E-05	7.18E-06	2.10E-08	2.77E-10

c MDS-L

d KG-1

e KG-1

f KG-1

g HEL
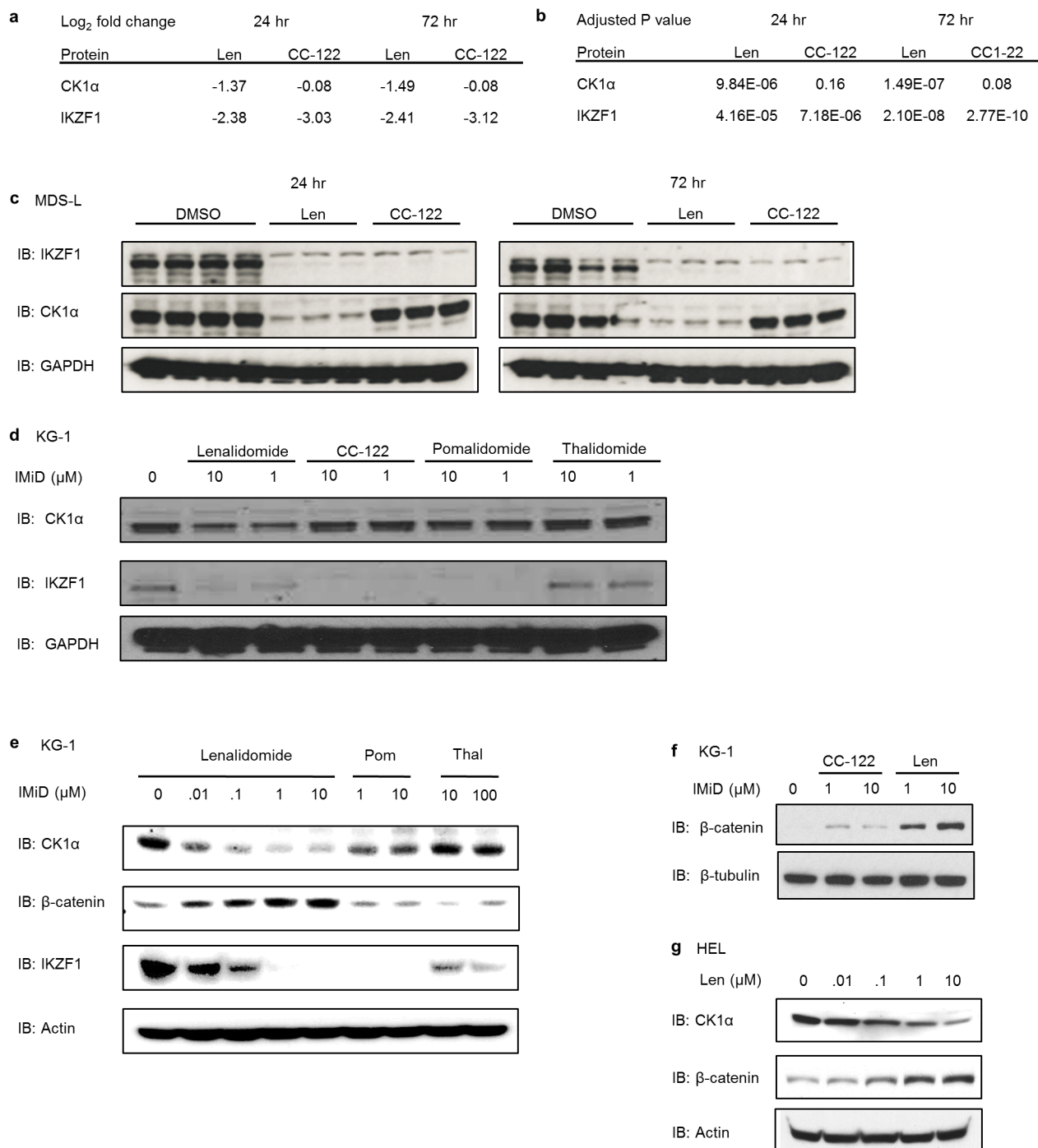
Extended Data Figure 8 | Comparison of the effects of thalidomide derivatives. **a**, Comparison of log₂ ratios for CK1α and IKZF1 in MDS-L cells after treatment with lenalidomide or CC-122 for 24 or 72 h assessed by tandem mass tag (TMT) quantitative proteomics. Analysis was performed with $n = 4$ for DMSO control and $n = 3$ for each drug treatment time point. **b**, Adjusted P values for CK1α and IKZF1 proteomic data in MDS-L cells. **c**, Western blot validation of IKZF1 and CK1α levels in DMSO ($n = 4$), lenalidomide ($n = 3$) and CC-122 ($n = 3$) treated samples used for MDS-L proteomic analysis.

d, Western blot validation of the effects of the different agents on CK1α and IKZF1 protein levels in KG-1 cells. **e**, Effect of lenalidomide, pomalidomide (Pom), and thalidomide (Thal) on protein levels of CK1α, β-catenin, and IKZF1 in KG-1 cells treated for 24 h with the indicated drug concentrations. **f**, Effect of CC-122 and lenalidomide on β-catenin protein levels in KG-1 cells after 72 h. **g**, Effect of lenalidomide on CK1α and β-catenin protein levels in HEL cells. Data are representative of two (**e**, **g**) or three (**c**, **d**) independent experiments. Uncropped blots are shown in Supplementary Fig. 1.

Extended Data Table 1 | Statistically significant SILAC results with 1 μM lenalidomide

a

K-ε-GG site	Replicate 1	Replicate 2	Average Log$_2$ fold change	adj.P.Val	Direction of change
IKZF1	2.78	3.18	2.98	7.23E-06	up
IKZF1	1.73	2.17	1.95	0.000497	up
MARCH8	1.58	0.92	1.25	0.036986	up
CK1α	1.06	1.08	1.07	0.035843	up
CRBN	-1.07	-1.21	-1.14	0.026006	down
RNF166	-1.49	-1.37	-1.43	0.003102	down
RNF166	-1.58	-1.39	-1.48	0.003102	down

b

Protein	Replicate 1	Replicate 2	Average Log$_2$ fold change	adj.P.Val	Direction of change
ZNF692	-1.89	-2.20	-2.05	0.013806	down
IKZF1	-1.62	-1.54	-1.58	0.005638	down
CK1α	-1.59	-1.53	-1.56	0.005638	down
RNF166	-1.41	-1.64	-1.52	0.015257	down
ZFP91	-0.69	-0.69	-0.69	0.047677	down
LEMD3	-0.66	-0.68	-0.67	0.047677	down
NRM	-0.64	-0.68	-0.66	0.047677	down
LBR	-0.67	-0.65	-0.66	0.047677	down
UNC84A.SUN1	-0.68	-0.64	-0.66	0.047677	down
C12orf57	0.66	0.70	0.68	0.047677	up

a, List of significantly regulated K-ε-GG sites with 1 μM lenalidomide vs DMSO. *P* value is adjusted as described in the methods section. **b,** List of significantly regulated proteins with 1 μM lenalidomide vs. DMSO. *P* value is adjusted as described in the methods section. Average log$_2$ fold change of two biological replicates.

Extended Data Table 2 | Characteristics of the patient samples from the AML-001 trial used in Extended Data Fig. 2b

Coded Patient Number (source of cells)	Sex	Primary Race	AML Classification	Peripheral Blood Blast Count	Age at Randomization	Prior MDS History?	MDS Primary Or Secondary	Study Arm Randomized To	Cycle 1 Dosing
1 (BMMC)	Male	White	AML not otherwise specified	>=1X10^9/L	71	Yes	Primary	Lenalidomide	50 mg daily, except drug withheld days 4-12
2 (BMMC)	Male	White	AML with myelodysplasia-related changes	>=1X10^9/L	80	Yes	Primary	Lenalidomide	50 mg daily, except drug withheld days 3-6 and 24-28
3 (PBMC)	Male	Asian	AML with myelodysplasia-related changes	<1X10^9/L	75	No		Lenalidomide	50 mg daily
4 (PBMC)	Male	White	AML with myelodysplasia-related changes	>=1X10^9/L	81	Yes	Primary	Lenalidomide	50 mg daily

BMMC: Bone marrow mononuclear cells
PBMC: Peripheral blood mononuclear cells

LETTER

A very luminous magnetar-powered supernova associated with an ultra-long γ-ray burst

Jochen Greiner[1,2], Paolo A. Mazzali[3,4], D. Alexander Kann[1,2,5], Thomas Krühler[6], Elena Pian[7,8], Simon Prentice[3], Felipe Olivares E.[9], Andrea Rossi[5,7], Sylvio Klose[5], Stefan Taubenberger[4,10], Fabian Knust[1], Paulo M. J. Afonso[11], Chris Ashall[3], Jan Bolmer[1,12], Corentin Delvaux[1], Roland Diehl[1], Jonathan Elliott[1,13], Robert Filgas[14], Johan P. U. Fynbo[15], John F. Graham[1], Ana Nicuesa Guelbenzu[5], Shiho Kobayashi[3], Giorgos Leloudas[15,16], Sandra Savaglio[1,17], Patricia Schady[1], Sebastian Schmidl[5], Tassilo Schweyer[1,12], Vladimir Sudilovsky[1,13], Mohit Tanga[1], Adria C. Updike[18], Hendrik van Eerten[1] & Karla Varela[1]

A new class of ultra-long-duration (more than 10,000 seconds) γ-ray bursts has recently been suggested[1–3]. They may originate in the explosion of stars with much larger radii than those producing normal long-duration γ-ray bursts[3,4] or in the tidal disruption of a star[3]. No clear supernova has yet been associated with an ultra-long-duration γ-ray burst. Here we report that a supernova (SN 2011kl) was associated with the ultra-long-duration γ-ray burst GRB 111209A, at a redshift z of 0.677. This supernova is more than three times more luminous than type Ic supernovae associated with long-duration γ-ray bursts[5–7], and its spectrum is distinctly different. The slope of the continuum resembles those of super-luminous supernovae[8,9], but extends further down into the rest-frame ultraviolet implying a low metal content. The light curve evolves much more rapidly than those of super-luminous supernovae. This combination of high luminosity and low metal-line opacity cannot be reconciled with typical type Ic supernovae, but can be reproduced by a model where extra energy is injected by a strongly magnetized neutron star (a magnetar), which has also been proposed as the explanation for super-luminous supernovae[10].

GRB 111209A was detected by the Swift satellite at 07:12 UT on 9 December 2011. The X-ray and optical counterparts were discovered within minutes[11]. The extraordinarily long duration of GRB 111209A was revealed by the continuous coverage provided by the Konus detector on the WIND spacecraft[12], extending from ~5,400 s before to ~10,000 s after the Swift trigger. The GRB occurred at a redshift of $z = 0.677$, as determined from afterglow spectroscopy[3]. Its integrated equivalent isotropic energy output, $E_{iso} = (5.7 \pm 0.7) \times 10^{53}$ erg (ref. 12), lies at the bright end of the distribution of long-duration GRBs.

The afterglow of GRB 111209A was observed over a period of about 70 days with the seven-channel optical/near-infrared imager GROND[13]. Starting around day 15, the optical light curve deviated from the earlier afterglow power-law decay (Fig. 1). The light curve remained essentially flat between days 15 and 30, and then started to decay again, approaching the host-galaxy level. After subtracting the afterglow and the well-modelled host galaxy emission (Methods, first three sections), the excess emission is well constrained between rest-frame days (that is, observed days divided by $(1 + z)$) 6 and 43 after the GRB (Fig. 2). This excess emission (Table 1) is very similar in shape to other GRB-related supernovae, but reaches a bolometric peak luminosity of $2.8^{+1.2}_{-1.0} \times 10^{43}$ erg s^{-1} (corresponding to a bolometric magnitude $M_{bol} = -20.0$ mag) at 14 rest-frame days, a factor of three times higher than the brightest known GRB-associated supernova (Fig. 2).

A spectrum was taken with the X-shooter instrument on the Very Large Telescope (ESO) near the peak of the excess emission[3] (29 December 2011), 11.8 rest-frame days after the GRB. The afterglow and the (minimal) host contribution were subtracted (Methods section 'The host galaxy') and the resulting spectrum is shown in Fig. 3 (blue line). The strong similarity of the evolution in time and colour to GRB-associated supernovae, together with the spectral shape of the excess emission, leads us to conclude that this emission is caused by a supernova, designated SN 2011kl, associated with GRB 111209A.

Canonical long-duration GRBs are generally accepted to be linked to the core collapse of massive stars stripped of their outer H and He envelopes[5–7], since every spectroscopically confirmed supernova associated with a GRB has been a broad-lined type Ic so far. Although the spectrum of SN 2011kl associated with the ultra-long GRB 111209A also shows no H or He, it is substantially different from classical GRB-associated supernovae. It is surprisingly featureless on the long-wavelength side ('redwards') of 300 nm, lacking the undulations from spectral line blends typical of broad-lined type Ic supernovae associated with GRBs[5–7], and it does not drop in the 300–400 nm (rest-frame) region (Fig. 3), suggesting a very low metal abundance. Applying standard parametrized supernova light-curve fits (Methods section 'Radioactivity cannot power the supernova peak'), we derive an ejecta mass $M_{ej} = 3 \pm 1\ M_\odot$ and a ^{56}Ni mass of $1.0 \pm 0.1\ M_\odot$, which implies a very high ^{56}Ni/M_{ej} ratio ($M_\odot$, solar mass). This large ^{56}Ni mass is not compatible with the spectrum, suggesting that ^{56}Ni is not responsible for the luminosity, unlike canonical stripped-envelope supernovae (Methods section 'Radioactivity cannot power the supernova peak').

Various models have been suggested to explain the ultra-long duration of GRB 111209A and other ultra-long GRBs, but the otherwise inconspicuous spectral and timing properties of both the prompt and afterglow emission as well as the host properties provided no obvious clues[1–4,14–16]. With the detection of a supernova associated with the ultra-long GRB 111209A, we can immediately discard a tidal disruption interpretation[3]. Known supernovae from blue supergiants show hydrogen in their spectra and substantially different light-curve

[1]Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstrasse 1, 85748 Garching, Germany. [2]Excellence Cluster Universe, Technische Universität München, Boltzmannstrasse 2, 85748 Garching, Germany. [3]Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Browlow Hill, Liverpool L3 5RF, UK. [4]Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85748 Garching, Germany. [5]Thüringer Landessternwarte Tautenburg, Sternwarte 5, 07778 Tautenburg, Germany. [6]European Southern Observatory, Alonso de Córdova 3107, Vitacura, Casilla 19001, Santiago 19, Chile. [7]INAF, Institute of Space Astrophysics and Cosmic Physics, via P. Gobetti 101, 40129 Bologna, Italy. [8]Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy. [9]Departamento de Ciencias Fisicas, Universidad Andres Bello, Avenida Republica 252, Santiago, Chile. [10]European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany. [11]American River College, Physics and Astronomy Department, 4700 College Oak Drive, Sacramento, California 95841, USA. [12]Technische Universität München, Physik Department, James-Franck-Strasse, 85748 Garching, Germany. [13]Astrophysics Data System, Harvard–Smithonian Center for Astrophysics, Garden Street 60, Cambridge, Massachusetts 02138, USA. [14]Institute of Experimental and Applied Physics, Czech Technical University in Prague, Horska 3a/22, 128 00 Prague 2, Czech Republic. [15]DARK Cosmology Center, Niels-Bohr-Institut, University of Copenhagen, Juliane Maries Vej 30, 2100 København, Denmark. [16]Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot 76100, Israel. [17]Universita della Calabria, 87036 Arcavacata di Rende, via P. Bucci, Italy. [18]Roger Williams University, 1 Old Ferry Road, Bristol, Rhode Island 02809, USA.
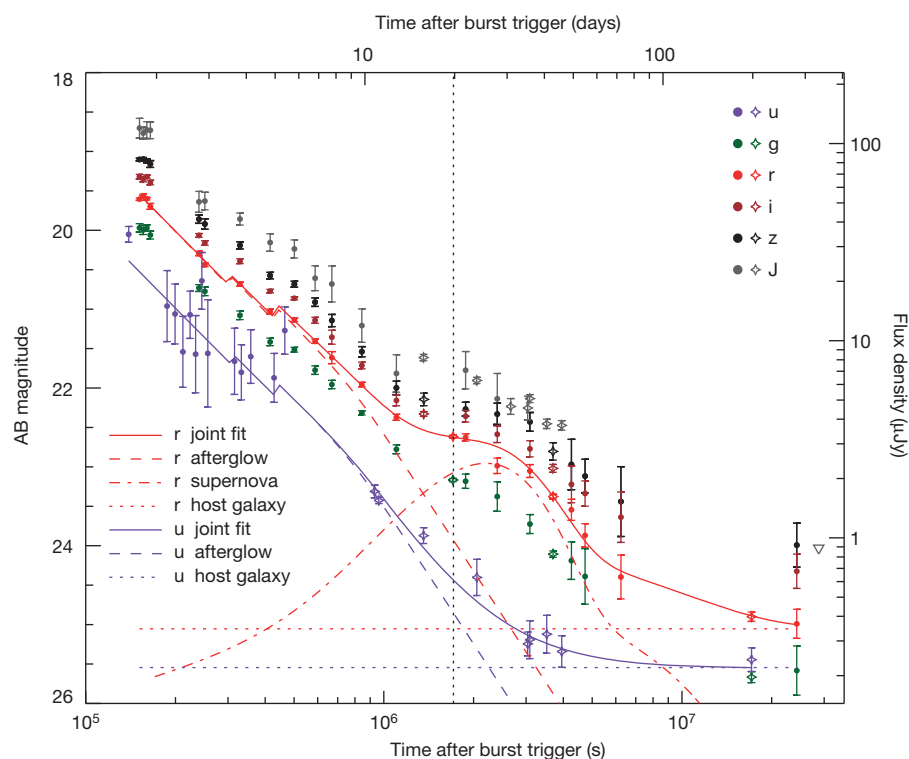
Figure 1 | Observed optical/near-infrared light curve of GRB 111209A. Data points (GROND data, filled symbols; other data, open symbols) show measured magnitudes. The fitted light curve (solid red line) is the sum of the afterglow of GRB 111209A modelled by a broken power law (dashed red line), the accompanying supernova SN 2011kl (dash-dotted red line) and the constant host galaxy emission (horizontal dotted red line). The u′-band data are well fitted without a supernova component, that is, the sum of only the afterglow and host (solid violet line). All measurements (error bars, 1σ uncertainty) are relative to the Swift trigger time and as observed, apart from the Vega-to-AB transformation for the J band. The vertical dotted line marks the time of the VLT/X-shooter spectrum.

properties[17], inconsistent with our observations, thus ruling out a blue supergiant progenitor[4]. Finally, additional emission from the interaction of the supernova ejecta with circumstellar material is also unlikely (Methods section 'Enhanced emission due to interaction with the circumburst medium?').

Our data suggest that SN 2011kl is intermediate between canonical overluminous GRB-associated supernovae and super-luminous supernovae (Fig. 3). The latter are a sub-class of supernovae that are

a factor of ~100 brighter than normal core-collapse supernovae, reaching a V-band magnitude $M_V \approx -21$ mag (refs 8, 9). They show slow rise times and late peak times (peak times about 20–100 days as compared to typically 9–18 days). Their spectra are characterized by a blue continuum with a distinctive "W"-shaped spectral feature often interpreted as O II lines[8]. A spinning-down magnetic neutron star is the favoured explanation for the energy input powering the light curve[10]. The comparison of SN 2011kl with super-luminous

Table 1 | AB magnitudes of SN 2011kl associated with GRB 111209A

Δt (s)	g′ mag	r′ mag	i′ mag	z′ mag	J mag
843, 664	$24.36^{+0.26}_{-0.21}$	$23.92^{+0.23}_{-0.19}$	$24.03^{+0.55}_{-0.38}$	$23.97^{+1.13}_{-0.57}$	
1,101,930	$24.17^{+0.29}_{-0.24}$	$23.66^{+0.16}_{-0.14}$	$23.80^{+0.44}_{-0.33}$	$23.83^{+0.75}_{-0.48}$	
1,358,649					22.38 ± 0.09
1,360,463			$23.28^{+0.12}_{-0.11}$		
1,361,742				$23.16^{+0.28}_{-0.25}$	
1,705,078	23.59 ± 0.04				
1,706,253		22.99 ± 0.04			
1,880,549	23.47 ± 0.15	22.90 ± 0.07	22.74 ± 0.13	$22.78^{+0.19}_{-0.18}$	$22.18^{+0.39}_{-0.35}$
2,049,952					22.30 ± 0.06
2,401,323	$23.53^{+0.28}_{-0.27}$	23.25 ± 0.15	22.90 ± 0.17	$22.67^{+0.23}_{-0.22}$	$22.54^{+0.53}_{-0.48}$
2,664,187					$22.62^{+0.16}_{-0.15}$
3,037,306					$22.58^{+0.22}_{-0.21}$
3,085,966					22.41 ± 0.07
3,090,966	$23.88^{+0.18}_{-0.17}$	23.21 ± 0.11	$23.05^{+0.17}_{-0.16}$	22.70 ± 0.19	
3,518,554					22.81 ± 0.09
3,692,304			23.35 ± 0.12		
3,693,574				$23.21^{+0.23}_{-0.22}$	
3,694,905	24.36 ± 0.07				
3,696,071		23.60 ± 0.05			
3,950,847					22.81 ± 0.09
4,258,444	$24.41^{+0.39}_{-0.37}$	23.80 ± 0.20	$23.63^{+0.42}_{-0.40}$	$23.44^{+0.62}_{-0.58}$	
4,732,196	$24.69^{+0.63}_{-0.58}$	$24.28^{+0.27}_{-0.26}$	$23.80^{+0.32}_{-0.31}$	$23.67^{+0.48}_{-0.46}$	
6,241,880		$25.26^{+0.84}_{-0.74}$	$24.29^{+0.78}_{-0.73}$	$24.27^{+1.57}_{-1.34}$	

The data are corrected for the GRB afterglow and host-galaxy contributions, as well as Galactic foreground and rest-frame extinction. Errors are at the 1σ confidence level and include error propagation from the afterglow and host subtraction. The first column (Δt) is the time after the GRB in the observer frame. The magnitudes without contemporaneous g′, r′, i′, z′ magnitudes are taken from ref. 3.

supernovae is motivated by two observational facts: (1) the spectrum is a blue continuum, extending far into the rest-frame ultraviolet, and (2) the peak luminosity is intermediate between GRB-associated supernovae and super-luminous supernovae. Our interpretation is motivated by the failure of both the collapsar and the standard fall-back accretion scenarios, because in these cases the engine quickly runs out of mass for any reasonable accretion rate and mass reservoir, and thus is unlikely to be able to power an ultra-long GRB.

We could reproduce the spectrum of SN 2011kl using a radiation transport code[18,19] and a radial (r) density (ρ) profile where $\rho \propto r^{-7}$, which is typical of the outer layers of supernova explosions. The ultraviolet emission is significantly depressed relative to a blackbody, but much less depressed than in the spectra of GRB-associated supernovae, indicating a lower metal content (consistent with 1/4 of the solar metallicity). The spectrum appears rather featureless owing to line blending. This follows from the high photospheric velocity, $v_{ph} \approx 20{,}000$ km s^{-1} (Fig. 3). In contrast, super-luminous supernovae, which show more line features, have $v_{ph} \approx 10{,}000$ km s^{-1}. In the optical part of the spectrum, on the other hand, only a few very weak absorption lines are visible in our supernova spectrum. Our model only has ~0.4 $M_\odot$ of material above the photosphere. There is no evidence of freshly synthesized material mixed-in, unlike the case of GRB-associated supernovae. This supports the notion that the supernova light curve was not powered by Ni decay but rather by a magnetar.

The supernova spectrum can be reproduced without invoking interaction, and the low metal abundance suggests that it is unlikely that much Ni was produced. We therefore consider magneto-rotational energy input as the source of luminosity. Using a simple formalism[20] describing rotational energy loss via magnetic dipole radiation, and relating the spin-down rate to the effective radiative diffusion time, we can infer the magnetar's initial spin period, P_i, and magnetic dipole field strength, B, from the observed luminosity and time to light-curve peak, t_{peak}. The observed short t_{peak} (~14 rest-frame days) and the

moderate peak luminosity require a magnetar with initial spin period $P_i \approx 12$ ms for a magnetic field strength of $(6-9) \times 10^{14}$ G. Depending on the magnetic field that is assumed, calculated values of ejecta mass and kinetic energy are relatively uncertain, ranging between 2 and 3 $M_\odot$ and $(2-9) \times 10^{51}$ erg, respectively (Methods section 'Modelling'). These values are actually more typical of normal type Ib/c supernovae than of GRB-associated supernovae, including SN 2006aj, the first supernova identified as magnetar-powered[21]. The GRB energy can be reconciled with the maximum energy that can be extracted from a magnetar if the correction for collimation of the GRB jet is a factor of 1/50 or less, which is well within typical values for GRBs[22].

The idea of a magnetar as the inner engine powering GRB-associated supernovae[23,24], super-luminous supernovae[10], or even events like Swift 1644+57[25] (before consensus for this event favoured a relativistic tidal disruption), is not new. However, in all these cases the magnetar interpretation was one of several options providing reasonable fits to the data, never the only option. Also, the suggestion that all GRB-associated supernovae are magnetars[24] rather than collapsars, based on the clustering of the kinetic energy of the GRB-associated supernovae near 10^{52} erg, the rotational power of a millisecond neutron star, was only circumstantial evidence for the magnetar origin. The supernova SN 2011kl is clearly different from canonical GRB-associated supernovae, and requires (rather than only allows) a new explanation.
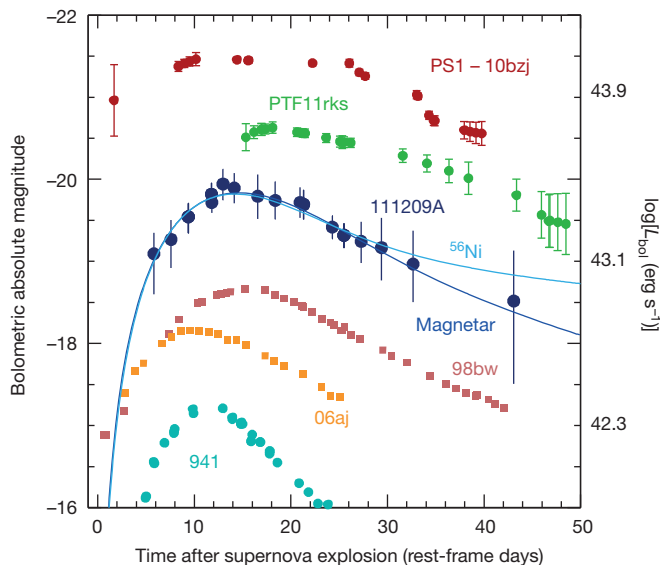
Figure 2 | Light curve of the supernova (SN 2011kl) linked with GRB 111209A and of other objects. Shown is the bolometric light curve of SN 2011kl, corresponding to 230–800 nm rest frame wavelengths (Methods section 'Observations and data analysis'), compared with those of GRB 980425/SN 1998bw[5], XRF 060218/SN 2006aj[21], the standard type Ic SN 1994I[26], and the super-luminous supernovae PTF11rks[27] and PS1-10bzj[28] (among the fastest-declining super-luminous supernovae known so far), all integrated over the same wavelength band with 1σ error bars. Solid lines show the best-fitting synthetic light curves computed with a magnetar injection model[20] (dark blue; Methods section 'Modelling') and ^{56}Ni powering (light blue; Methods section 'Radioactivity cannot power the supernova peak').
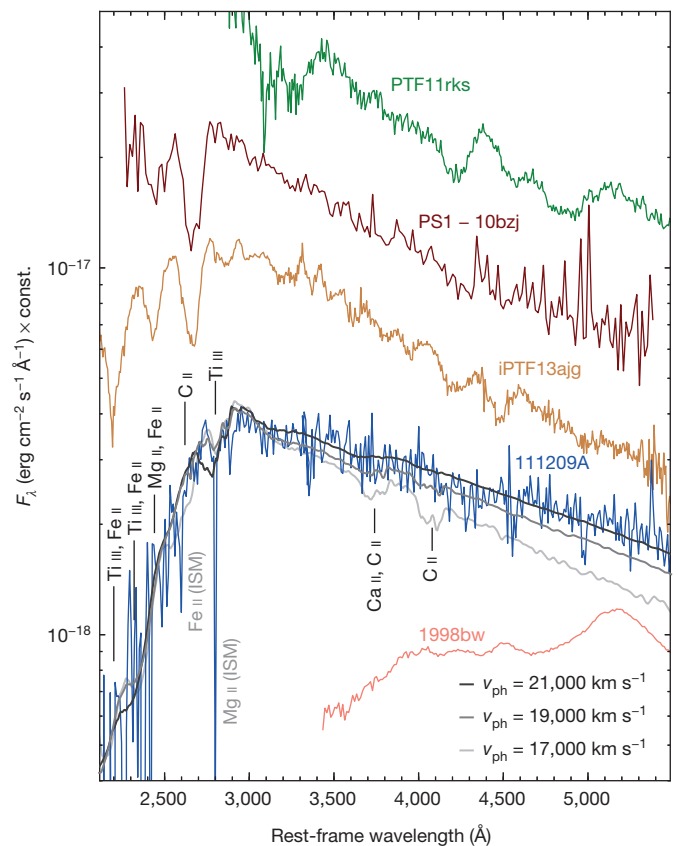
Figure 3 | Spectra comparison. The X-shooter spectrum of SN 2011kl, associated with GRB 111209A, taken on 29 December 2011 after GRB afterglow and host subtraction and moderate rebinning (Methods section 'Observations and data analysis'; Extended Data Fig. 2), with its flat shape and high ultraviolet flux, is distinctly different from the hitherto brightest known GRB-associated supernova 1998bw (red), but reminiscent of some super-luminous supernovae (top three curves)[28–30]. The three grey/black lines show synthetic spectra with different photospheric velocities (as labelled), demonstrating the minimum velocity required to broaden unseen absorption around 400 nm rest-frame (Ca II, C II), but at the same time explain the sharp cut-off below 280 nm rest-frame. The y scale is correct for SN 2011kl and SN 1998bw; all other spectra are shifted for display purposes.

The ultra-long duration of the prompt emission of GRB 111209A and the unusual supernova properties are probably related. We suggest that they are linked to the birth and subsequent action of a magnetar following the collapse of a massive star. The magnetar re-energizes the expanding ejecta and powers an over-luminous supernova. This particular supernova, SN 2011kl, was not quite as luminous as typical super-luminous supernovae, and it may represent a population of events that is not easily discovered by supernova searches but which may occur at a relatively high rate. This scenario offers a link between GRB-associated supernovae, ultra-long GRBs and super-luminous supernovae.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Gendre, B. *et al.* The ultra-long gamma-ray burst 111209A: the collapse of a blue supergiant? *Astrophys. J.* **766,** 30 (2013).
2. Stratta, G. *et al.* The ultra-long gamma-ray burst 111209A: II. Prompt to afterglow and afterglow properties. *Astrophys. J.* **779,** 66 (2013).
3. Levan, A. *et al.* A new population of ultra-long duration gamma-ray bursts. *Astrophys. J.* **781,** 13 (2014).
4. Nakauchi, D., Kashiyama, K., Suwa, Y. & Nakamura, T. Blue supergiant model for ultra-long gamma-ray bursts with superluminous-supernova-like bump. *Astrophys. J.* **778,** 67 (2013).
5. Galama, T. *et al.* An unusual supernova in the error box of the γ-ray burst of 25 April 1998. *Nature* **395,** 670–672 (1998).
6. Hjorth, J. *et al.* A very energetic supernova associated with the γ-ray burst of 29 March 2003. *Nature* **423,** 847–850 (2003).
7. Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591,** L17–L20 (2003).
8. Quimby, R. M. *et al.* Hydrogen-poor superluminous stellar explosions. *Nature* **474,** 487–489 (2011).
9. Gal-Yam, A. Luminous supernovae. *Science* **337,** 927–932 (2012).
10. Dessart, L., Hillier, D. J., Waldman, R., Livne, E. & Blondin, S. Superluminous supernovae: Ni power versus magnetar radiation. *Mon. Not. R. Astron. Soc.* **426,** L76–L80 (2012).
11. Hoversten, E. A. *et al.* GRB 111209A: Swift detection of a long burst with an optical counterpart. *GCN Circ.* **12632** (2011).
12. Golenetskii, S. *et al.* Konus-Wind observation of GRB 111209A. *GCN Circ.* **12663** (2011).
13. Greiner, J. *et al.* GROND — a 7-channel imager. *Publ. Astron. Soc. Pacif.* **120,** 405–424 (2008).
14. Virgili, F. J. *et al.* GRB 091024A and the nature of ultra-long gamma-ray bursts. *Astrophys. J.* **778,** 54 (2013).
15. Zhang, B.-B., Zhang, B., Murase, K., Connaughton, V. & Briggs, M. S. How long does a burst burst? *Astrophys. J.* **787,** 66 (2014).
16. Leloudas, G. *et al.* Spectroscopy of superluminous supernova host galaxies. A preference of hydrogen-poor events for extreme emission line galaxies. *Mon. Not. R. Astron. Soc.* **449,** 917–932 (2015).
17. Kleiser, I. K. W. *et al.* Peculiar Type II supernovae from blue supergiants. *Mon. Not. R. Astron. Soc.* **415,** 372–382 (2011).
18. Mazzali, P. A. & Lucy, L. B. The application of Monte Carlo methods to the synthesis of early-time supernovae spectra. *Astron. Astrophys.* **279,** 447–456 (1993).
19. Mazzali, P. A. Applications of an improved Monte Carlo code to the synthesis of early-time supernova spectra. *Astron. Astrophys.* **363,** 705–716 (2000).
20. Kasen, D. & Bildsten, L. Supernova light curves powered by young magnetars. *Astrophys. J.* **717,** 245–249 (2010).
21. Mazzali, P. A. *et al.* A neutron-star-driven X-ray flash associated with supernova SN 2006aj. *Nature* **442,** 1018–1020 (2006).
22. Ryan, G., van Eerten, H., MacFadyen, A. & Zhang, B.-B. Gamma-ray bursts are observed off-axis. *Astrophys. J.* **799,** A3 (2015).
23. Cano, Z. A trio of gamma-ray burst supernovae: GRB 120729A, GRB 130215A/SN 2013ez, and GRB 130831A/SN 2013fu. *Astron. Astrophys.* **568,** 19 (2014).
24. Mazzali, P. A. *et al.* An upper limit to the energy of gamma-ray bursts indicates that GRBs/SNe are powered by magnetars. *Mon. Not. R. Astron. Soc.* **443,** 67–71 (2014).
25. Quataert, E. & Kasen, D. Swift 1644+57: the longest gamma-ray burst? *Mon. Not. R. Astron. Soc.* **419,** L1–L5 (2012).
26. Sauer, D. N. *et al.* The properties of the 'standard' type Ic supernova 1994I from spectral models. *Mon. Not. R. Astron. Soc.* **369,** 1939–1948 (2006).
27. Inserra, C. Super-luminous Type Ic supernovae: catching a magnetar by the tail. *Astrophys. J.* **770,** 128 (2013).
28. Lunnan, R. *et al.* PS1–10bzj: a fast, hydrogen-poor superluminous supernova in a metal-poor host galaxy. *Astrophys. J.* **771,** 97 (2013).
29. Vreeswijk, P. M. *et al.* The hydrogen-poor superluminous supernova iPTF 13ajg and its host galaxy in absorption and emission. *Astrophys. J.* **797,** 24 (2014).
30. Yaron, O. & Gal-Yam, A. WISeREP — an interactive supernova data repository. *Publ. Astron. Soc. Pacif.* **124,** 668–681 (2012).

METHODS

Observations and data analysis. Simultaneous imaging in g′, r′, i′, z′, J, H, K$_s$ with the 7-channel imager GROND[13] was done on 16 epochs with logarithmic temporal spacing until 72 days after the GRB, when the nearby Sun prevented further observations, and a last epoch for host photometry was obtained 280 days after the GRB (Extended Data Table 1). GROND data have been reduced in the standard manner using pyraf/IRAF[31–33]. The optical imaging was calibrated against comparison stars obtained by observing a nearby SDSS field (immediately before the afterglow observation in the third night under photometric conditions) and calibrated against the primary SDSS[34] standard star network. The near-infrared data were calibrated against the 2MASS catalogue. This results in typical absolute accuracies of ±0.03 mag in g′, r′, i′, z′, and ±0.05 mag in J, H, K$_S$ (1σ errors are reported everywhere). All GROND measurements are listed in Extended Data Table 1, and the properties of the GRB afterglow proper, including the two kinks in the early afterglow light curve (Fig. 1) will be described in detail elsewhere (D.A.K. *et al.*, manuscript in preparation).

We have made use of two other sources of measurements: First, we add u-band observations obtained with Swift/UVOT (Extended Data Table 2). UVOT photometry was carried out on pipeline-processed sky images downloaded from the Swift data centre[35] following the standard UVOT procedure[36], and is fully compatible with earlier, independent publications of the UVOT data[2,3]. Second, we add selected complementary data[3], in particular (i) HST F336W/F125W data from 11.1 and 35.1 days after the GRB, respectively; (ii) two epochs of VLT/FORS2 g′, R$_C$, i′, z′, data during the supernova phase, which agree excellently with our data due to their use of our GROND calibration stars; (iii) a late-time Gemini-S u′-band observation (198 days after the GRB).

With the constant host galaxy contribution accurately determined at late times in u′, g′, r′, i′, z′, J (see Methods section 'The host galaxy' and Extended Data Fig. 4), the afterglow light curve shows clear evidence for a steeper afterglow decay at >10 days post-burst, particularly in the u′-band where there is essentially no contribution from the supernova (as evidenced by the spectrum) and which therefore can be used as a template for the pure afterglow contribution. We link the decay slopes for all filters to each other, so we use the same single fit parameter for all filters. This provides the two decay slopes $\alpha_1 = 1.55 \pm 0.01$ and $\alpha_2 = 2.33 \pm 0.16$, with a break time of $t_b = 9.12 \pm 0.47$ days. The u′-band fit is also shown in Fig. 1 to visualize the decomposition. Apart from our much larger data set provided by our GROND observations, the difference between our fit and the decomposition of ref. 3 is the fact that in the latter the host contribution in the redder bands at ~30–50 days was ignored (although this is noted in ref. 3).

In order to create the supernova light curve for each photometric band, we then subtracted both the afterglow contribution in that band based on the extrapolation of the afterglow light curve, and the host galaxy contribution based on its spectral energy distribution; see Methods section 'The host galaxy'. The error in the host galaxy subtraction is negligible as the host photometry is accurate to better than 10%, and the host contributes only 5–15% to the total light during the supernova bump. The error on the afterglow subtraction depends on whether or not the decay slope remained constant after the last secure measurement right before the onset of the supernova. The intrinsic GRB afterglow light curves at this late time are observed to only steepen, never flatten. Thus, our afterglow subtraction is conservative, and results in a lower limit for the supernova luminosity.

The quasi-bolometric light curve of SN 2011kl was constructed from GROND g′, r′, i′, z′, J photometry and the supplementary data from ref. 3 as follows. First, the individual filter bands have been extinction-corrected with $A_V^{Gal} = 0.06$ mag Galactic foreground[37], and rest-frame $A_V^{Host} = 0.12$ mag as derived from the GRB afterglow spectral energy distribution fitting. By deriving quadratic polynomials for sets of three consecutive filters (Simpson's rule), they were then combined to create a quasi-bolometric light curve.

The quadratic polynomials are then integrated over rest-frame wavelength from $3,860/(1 + z)$ Å (blue edge of the g′-band filter) to $13,560/(1 + z)$ Å (red edge of the J filter). The k-correction was computed from the spectral energy distribution. In order to transform the integrated flux into luminosity, we employed a luminosity distance of $d = 4,080$ Mpc, using concordance cosmology ($\Omega_\Lambda = 0.73$, $\Omega_m = 0.27$, and $H_0 = 71$ km s^{-1} Mpc^{-1}).

No correction for the contribution of the unobserved near-infrared part of the spectrum has been applied to SN 2011kl or SN 1998bw (Fig. 2), because this emission is usually sparsely sampled in wavelength and time, and thus is largely based on assumptions (and no data are available for the plotted super-luminous supernovae). For SN 2011kl we lack any rest-frame near-infrared measurements. We acknowledge that therefore the bolometric luminosity might be underestimated by 5–30%. Other than that, all bolometric light curves shown in Fig. 2 are integrated over the same wavelength band (except for the ultraviolet band, which contributes less than a few percent at and after maximum). The super-luminous supernovae light curves are plotted according to the observational constraints of

their maxima, that is, g-band peak at 16.8 days rest-frame for PTF11rks[27] and using the first measurement 17.5 days before maximum as lower limit for PS1-10bzj[28].

The VLT/X-shooter[38] spectrum, taken on 29 December 2011 (19.8 days after the GRB, 11.8 rest-frame days, and 2 days before the supernova maximum), has been reduced with the ESO X-shooter pipeline v2.2.0, in particular for flat-fielding, order tracing, rectification and initial wavelength calibration with an arc lamp. During rectification, a dispersion of 0.4 Å per pixel has been used in the UVB/VIS arm, minimizing correlated noise but maintaining sufficient spectral resolution for resolving lines down to ~50 km s^{-1}, that is, a velocity dispersion of 20 km s^{-1}. Our own software is used for bad-pixel and cosmic-ray rejection, as well as sky-subtraction and frame shifting and adding[39]. Optimal extraction is applied to the resulting two-dimensional frames, and the one-dimensional spectrum is finally flux calibrated separately for each arm against the GROND photometry. Spectral binning has no effect on the steepness of the slope (Extended Data Fig. 1). The NIR arm does not contain any useful signal, nor do the two HST grism spectra[3] (Extended Data Fig. 2).

The observed spectrum is the sum of light from the GRB afterglow, the GRB host galaxy, and the supernova SN 2011kl. After correcting for $A_V^{Gal} = 0.06$ mag Galactic foreground[37] extinction, we corrected for the contribution of the host galaxy using a template fit (Methods section 'The host galaxy') on the host photometry (including the J-band measurement of ref. 3), and subtracted the afterglow based on the extrapolation of the g′, r′, i′, z′ GROND light curves to the time of the X-shooter observation. After conversion to the rest-frame, we corrected for intrinsic reddening of $E(B - V) = 0.04 \pm 0.01$ mag derived from the GROND afterglow SED fitting (see Extended Data Fig. 3 for the effect of each of these steps).

Association of GRB afterglow, supernova, and host galaxy. We detect narrow absorption lines of Mg II(λ2796, λ2803), Mg I(λ2852) and Fe II(λ2344, λ2374, λ2382, λ2586, λ2600) in the SN 2011kl spectrum. No change in equivalent widths and redshift is apparent when compared to the afterglow spectrum[3,39] taken 0.75 days after the GRB. Moreover, these equivalent widths are typical of those seen from host galaxies of bright long-duration GRBs. This relates the supernova to the same host galaxy as GRB 111209A.

No offset is measurable in GROND images between GRB afterglow and supernova (δRA < 0.032 arcsec, δDec. < 0.019 arcsec), which implies that the two events are co-spatial within <200 pc.

The host galaxy. During the late-epoch GROND observation the host galaxy is clearly detected in g′, r′, i′, z′ in the 3−5σ range (last entry in Extended Data Table 1). We add HST F336W and Gemini from ref. 3. Noting that the supernova does not contribute significantly any more during these late epochs (with expected AB magnitudes g′ ≈ 28.5, r′ ≈ 28.0, i′ ≈ 27.5, z′ ≈ 27.2 mag), we employ LePHARE[40] and use the best-fit model (a low-mass, star-forming galaxy) as a template for the host subtraction (see Extended Data Figs 3 and 4). Inferences on the physical properties of the host from this fitting will be published elsewhere (D.A.K. *et al.*, manuscript in preparation) and absorption/emission line information from the optical/near-infrared X-shooter spectra are given in ref. 39. We note though that the low metal content seen in the supernova spectrum is in accord with the very low host galaxy metallicity (10–40%), which is somewhat unusual for such a low-redshift object but commonly seen in super-luminous supernova hosts.

Radioactivity cannot power the supernova peak. Modelling the bolometric light curve according to the standard scheme of ^{56}Ni powering[41] and augmented by Co decay[42], an ejecta mass of 3.2 ± 0.5 $M_\odot$ and a ^{56}Ni mass of 1.0 ± 0.1 $M_\odot$ are derived (we used $v_{ph} = 20,000$ km s^{-1}, and a grey opacity of 0.07 ± 0.01 cm^2 g^{-1}, constant in time). The derived ^{56}Ni mass is anomalously large for type Ib/c supernovae, including GRB-associated supernovae[43]. Such a large ^{56}Ni mass is difficult to reconcile with the very low opacity in the blue part of the spectrum. The continuum flux keeps rising down to 300 nm rest-frame without any sign of suppression implying very low metal line opacity. Also, the ejected mass of ~3 $M_\odot$ as deduced from the light curve width is not consistent with the large ^{56}Ni mass.

While it has been suggested that part of the ^{56}Ni could be synthesized in the accretion disk[44], this is unlikely to proceed at the rate needed in our case. Recent numerical simulations show that for a wide range of progenitor masses (13–40 $M_\odot$), initial surface rotational velocities, metallicities and explosion energies, the required disk mass of more than 1 $M_\odot$ (corresponding to ~0.2 $M_\odot$ ^{56}Ni) is difficult to achieve[45], for both cases of compact objects: (i) in the case of heavy fallback, leading to the collapse of the central object into a black hole, the explosion energy is required to be small (few × 10^{48} erg), and more importantly, the disk forms only after a few months due to the large fallback time (~10^6 s); (ii) in the case of little fallback, leaving a neutron star behind, only fine-tuned conditions produce fallback disks at all, and these then have lifetimes of at most several hundred seconds.

Thus, a different mechanism must power the supernova light curve during the first ~40 days (rest frame).

Enhanced emission due to interaction with the circumburst medium?. Given the large luminosity, we considered additional emission from the interaction of the supernova ejecta with the circumstellar medium as an alternative possibility. In that case, one may expect narrow Balmer emission lines. While we detect very narrow ($\sigma = 35$ km s^{-1}) Hα, Hβ and [O III] lines in emission, the Balmer fluxes are compatible with the forbidden line flux and with an origin from the global low (0.02 $M_\odot$ yr^{-1}) star formation rate in this low-metallicity (10–40% solar) host galaxy[39]. On the other hand, if the progenitor star was heavily stripped, no circumstellar H may be present. Another, more serious constraint is the very blue supernova spectrum, which would require a very low density to minimize extinction (though dust may be destroyed by the initial GRB and supernova light). This may be at odds with the requirement that the density is high enough to generate the few 10^{43} erg s^{-1} of radiative luminosity observed around the peak.

Modelling. We have been able to reproduce the spectrum of SN 2011kl using a radiation transport code[18,19] and a radial (r) density (ρ) profile where $\rho \propto r^{-7}$, which is typical of the outer layers of supernova explosions. The spectra appear rather featureless but this does not mean that there is no absorption: the ultraviolet is significantly depressed relative to a blackbody. However, it is much less depressed than in the spectra of GRB-associated supernovae, indicating a lower metal content. Many metal lines are active in the ultraviolet (Fe, Co, Ti, Cr). The smooth appearance of the ultraviolet spectrum is the result of the blending of hundreds of lines caused by the large range of wavelengths over which lines are active (line blanketing). Indeed, the photospheric velocity (and density) determines the degree of line blending. We used here photospheric velocities of $v_{ph} \approx$ 20,000 km s^{-1} (grey/black lines in Fig. 3), and can see increasingly featureless spectra as v_{ph} increases and lines are active at higher velocities (larger blueshift), demonstrating the minimum velocity required to broaden unseen absorption around 400 nm rest-frame (Ca II, C II) and at the same time explain the sharp cut-off below 280 nm rest-frame. The strongest lines that shape this strong blue cut-off are labelled in black (grey 'ISM' labels mark Mg II/Fe II absorption lines in the host galaxy). Most of these are blended and do not stand out as individual features, unlike in classical super-luminous supernovae which have $v_{ph} \approx 10,000$ km s^{-1}. In the optical, on the other hand, only a few very weak absorption lines are visible in our supernova spectrum. These are due to Ca II and C II lines. O II lines are not detected, and would require large departures from thermal equilibrium because of the very high ionization/excitation potential of their lower levels (20–30 eV). This suggests the presence of X-rays in super-luminous supernovae, probably produced by shocks. Our model only has ~0.4 $M_\odot$ of material above the photosphere. The metal content is quite low. It is consistent with 1/4 of the solar metallicity, which could be the metallicity of the star whose explosion caused the GRB and the supernova, and there is no evidence of freshly synthesized material mixed-in, unlike in GRB-associated supernovae. This supports the notion that the supernova light curve was not powered by Ni decay but rather by a magnetar. Figure 3 shows this model with three different photospheric velocities overplotted on the X-shooter spectrum.

The spectrum can be reproduced without invoking interaction, but the metal abundance is so low that it is unlikely that much ^{56}Ni has been produced. We therefore consider magneto-rotational energy input as the source of luminosity. Depending on the relative strength of magnetar and radioactive decay energy deposition, different peak luminosities as well as rise and decay times can be obtained[20]. One particularly pleasant feature of the magnetar mechanism is that it does not necessarily suffer from strong line blanketing, thus providing a more natural explanation for the observed spectrum.

Using a simple formalism describing rotational energy loss via magnetic dipole radiation and relating the spin-down rate to the effective radiative diffusion time, one can infer the magnetar's initial spin period P_i and magnetic dipole field strength from the observed luminosity and time to light curve peak t_{peak}. One million combinations of the parameters P_i, B, M_{ej} and E_K were sampled and ranked according to the goodness of fit relative to the data. All best solutions cluster at $P_i = 12.2 \pm 0.3$ ms and have $B = (7.5 \pm 1.5) \times 10^{14}$ G, required by the observed short t_{peak} (~14 rest-frame days) and the moderate (for a magnetar) peak luminosity. Under the assumption that the magnetar is the sole contributor to the kinetic energy and to the light curve, a larger energy would be required from the magnetar, leading to $P_i \approx 3$ ms and $B \approx 4 \times 10^{14}$ G. The mass and energy of the ejecta are less well determined, as they depend on the energy injection by the magnetar, and also due to the unknown distribution of mass in velocity space below the photosphere. We find a rather low ejected mass $M_{ej} = 2.4 \pm 0.7$ $M_\odot$, and energy $E_K = (5.5 \pm 3.3) \times 10^{51}$ erg. Different photospheric velocities of, for example, 10,000, 15,000 and 20,000 km s^{-1} lead to different ejecta masses of 1.1, 1.7 and 2.2 $M_\odot$, but produce indistinguishable light curves with $M_{Ni} = 1.1 \pm 0.1$ $M_\odot$. Note though that not every combination of P_i, M_{ej} and E_K yields similar results. The GRB energy can be reconciled with the maximum energy that can be extracted from a magnetar if the correction for collimation of the GRB jet is a factor of 1/50 or less, which is well within typical values for GRBs[22].

Code availability. The code used in refs 18, 19 is available on request from mazzali@mpa-garching.mpg.de.

31. Tody, D. in *Astronomical Data Analysis Software and Systems II* (eds Hanisch, R. J., Brissenden, R. J. V. & Barnes, J.) 173 (ASP Conf. Ser. 52, 1993).
32. Krühler, T. *et al.* The 2175 Å feature in a gamma-ray burst afterglow at redshift 2.45. *Astrophys. J.* **685,** 376–383 (2008).
33. Küpcü Yoldaş, A. *et al.* First results of GROND. *AIP Conf. Proc.* **1 000,** 227–231 (2008).
34. SDSS. http://www.sdss.org/dr3/algorithms/fluxcal.html (accessed 21 October 2014).
35. Swift: catching gamma-ray bursts on the fly. http://www.swift.ac.uk/swift_portal (2014).
36. Poole, T. S. *et al.* Photometric calibration of the Swift ultraviolet/optical telescope. *Mon. Not. R. Astron. Soc.* **383,** 627–645 (2008).
37. Schlegel, D., Finkbeiner, D. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500,** 525–553 (1998).
38. Vernet, J. *et al.* X-shooter, the new wide band intermediate resolution spectrograph at the ESO Very Large Telescope. *Astron. Astrophys.* **536,** A105 (2011).
39. Krühler, T. *et al.* GRB hosts through cosmic time: an emission-line survey of 96 gamma-ray burst selected galaxies between 0.1 < z < 3.6. *Astron. Astrophys.* (in the press); preprint at http://arXiv.org/abs/1505.06743 (2015).
40. Arnouts, S. & Le Phare, O. Le PHARE: Photometric analysis for redshift estimate. http://www.cfht.hawaii.edu/~arnouts/LEPHARE (2014).
41. Arnett, W. D. Type I supernovae. I — Analytic solutions for the early part of the light curve. *Astrophys. J.* **253,** 785–797 (1982).
42. Valenti, S. *et al.* The broad-lined Type Ic supernova 2003jd. *Mon. Not. R. Astron. Soc.* **383,** 1485–1500 (2008).
43. Mazzali, P. A. *et al.* The very energetic, broad-lined Type Ic supernova 2010ah (PTF10bzf) in the context of GRB/SNe. *Mon. Not. R. Astron. Soc.* **432,** 2463–2473 (2013).
44. Surman, R., McLaughlin, G. C. & Hix, W. R. Nucleosynthesis in the outflow from gamma-ray burst accretion disks. *Astrophys. J.* **643,** 1057–1064 (2006).
45. Perna, R., Duffell, P., Cantiello, M. & MacFadyen, A. I. The fate of fallback matter around newly born compact objects. *Astrophys. J.* **781,** 119 (2014).
46. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345,** 245–256 (1989).
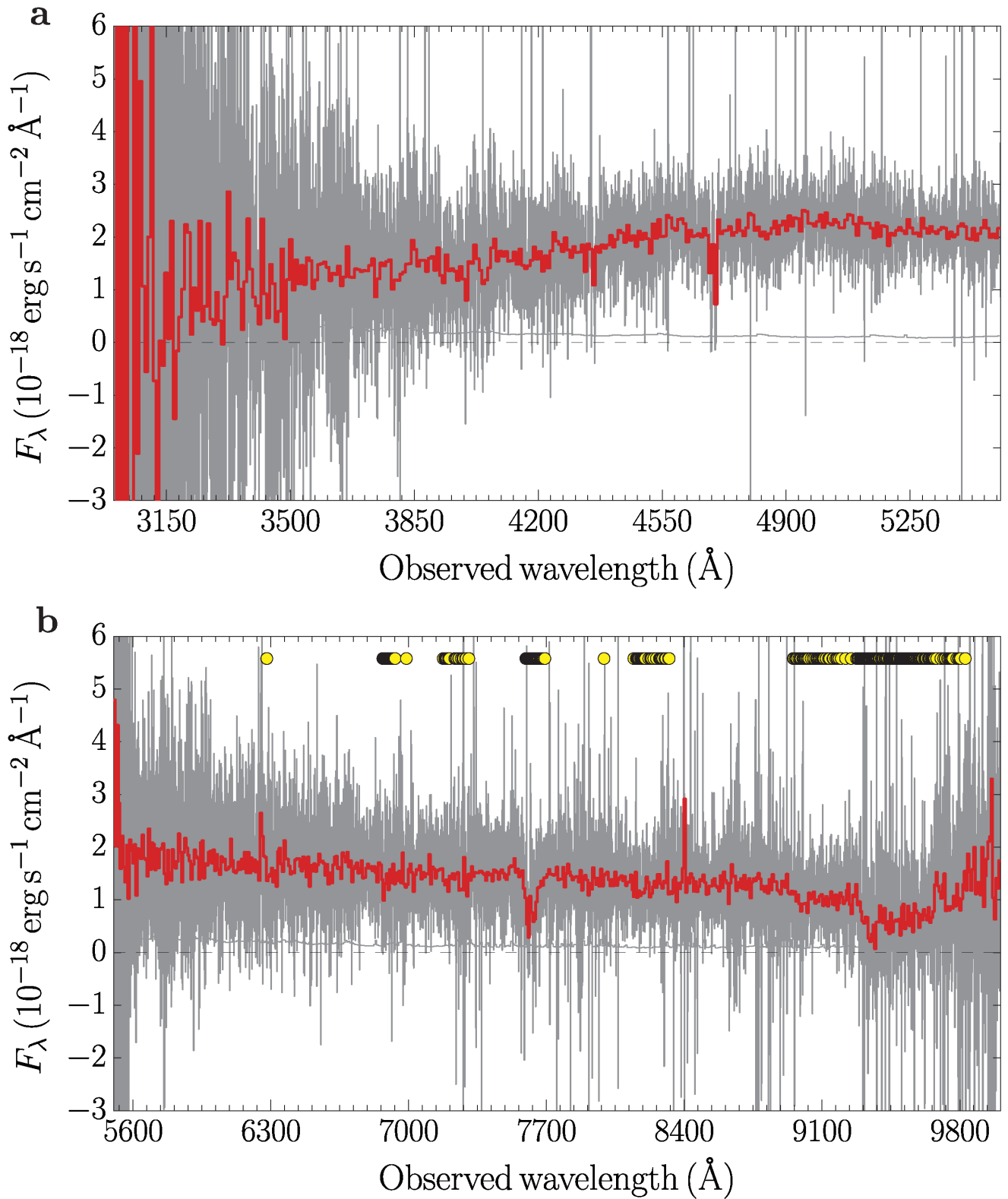
Extended Data Figure 1 | Binning has no effect on spectral slope.
Original X-shooter spectrum in the UVB (**a**) and VIS (**b**) arms shown in grey (0.4 Å per pixel; before host and afterglow subtraction), with the re-binned (factor of 20) spectrum overplotted in black. The binning does not change the steepness of the spectrum, in particular not at the blue end. Yellow circles denote positions of atmospheric absorption lines.
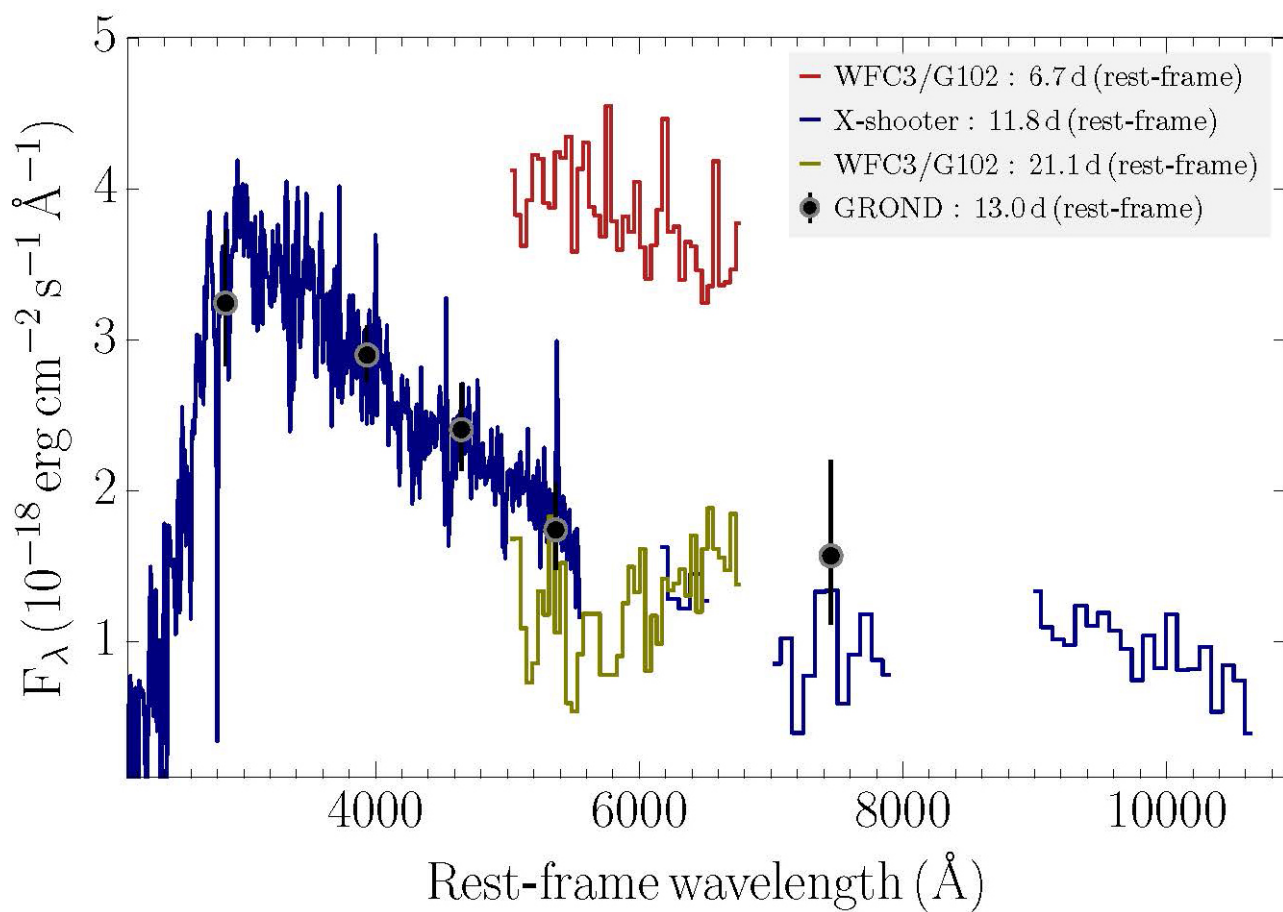
Extended Data Figure 2 | Long-wavelength spectra. Full X-shooter spectrum near maximum light of SN 2011kl, as well as two HST grism spectra taken one week before and after the supernova maximum (both taken from ref. 3). Above 500 nm rest-frame, none contain any informative absorption lines (all absorption structures seen are from the Earth's atmosphere).
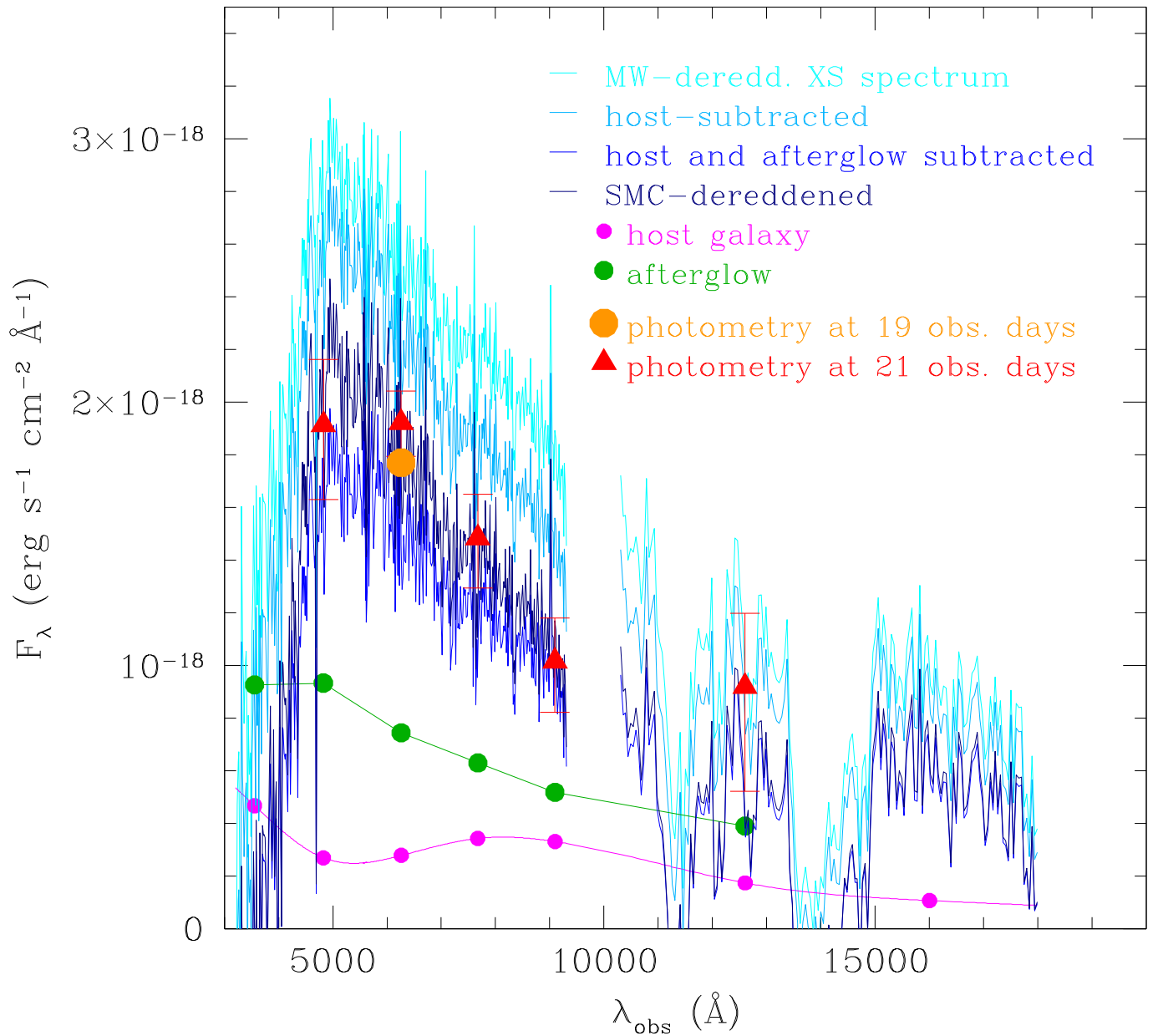
Extended Data Figure 3 | Step-by-step corrections of the supernova spectrum. Sequence of analysis steps for the X-shooter spectrum; from the observed spectrum corrected only for galactic foreground (top, very light blue), through host subtraction (light blue) and afterglow+host subtraction (blue) to local host (SMC-like) dereddened (very dark blue). The break at 500 nm observer-frame (300 nm rest-frame) and the steep slope towards the ultraviolet are inherent to the raw spectrum, not a result of afterglow or host subtraction. The coloured data points are the photometric observations in the individual UVOT+GROND+Gemini filters.
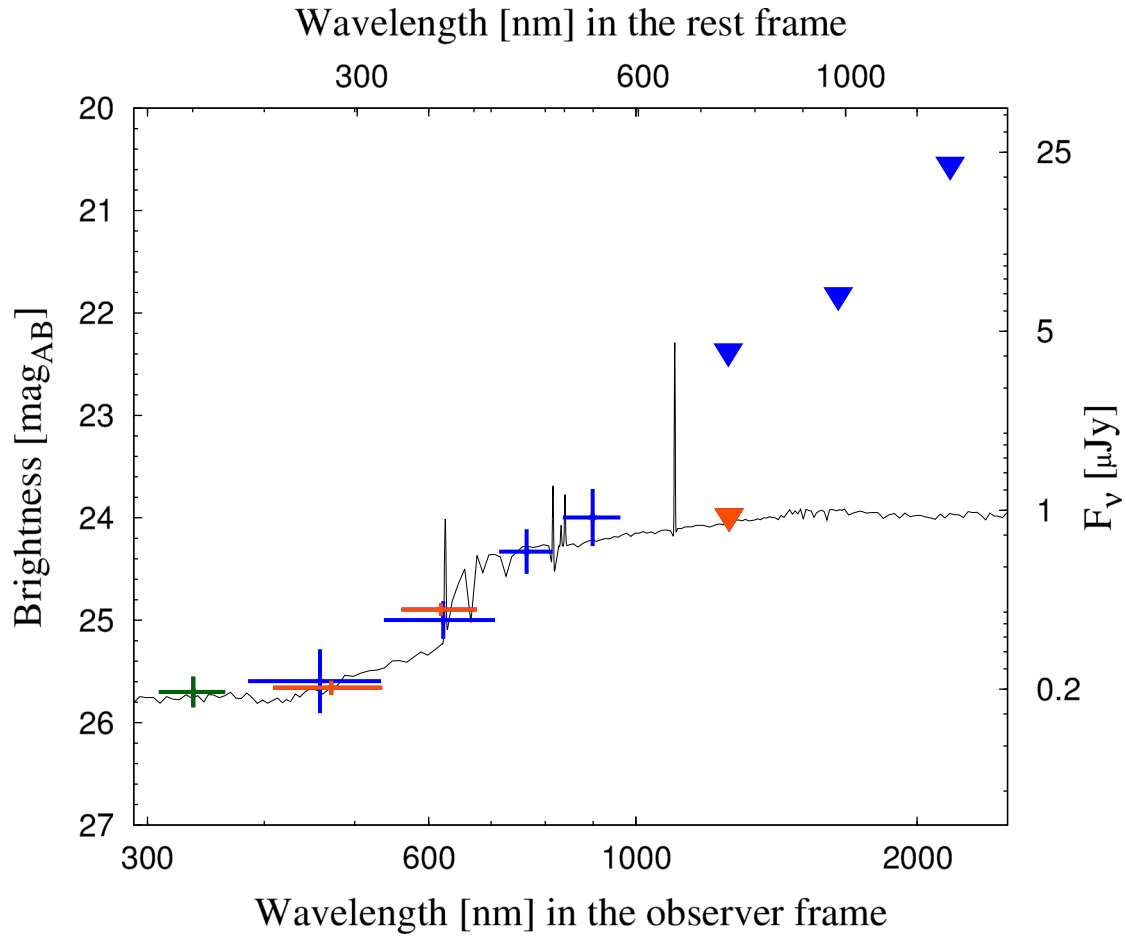
Extended Data Figure 4 | Observed spectral energy distribution of the host galaxy of GRB 111209A. Plotted in blue are GROND g′, r′, i′, z′ detections with 1σ errors (crosses) and GROND J, H, K$_S$ upper limits (3σ; triangles) of the host galaxy of GRB 111209A. Data taken from ref. 3 are F336W (green), Gemini g′, r′ detections (red crosses) and the J-band upper limit (red triangle). The best-fit LePHARE template of a low-mass, low-extinction, young star-forming galaxy is shown, which is very typical for GRB host galaxies.

Extended Data Table 1 | GROND observations of the afterglow, supernova and host of GRB 111209A

Δt (ks)	T_{VIS} (s)	g′ (mag)	r′ (mag)	i′ (mag)	z′ (mag)	T_{NIR} (s)	J (mag)	H (mag)	K_s (mag)
151.49	460	20.05± 0.05	19.66± 0.02	19.36± 0.03	19.13± 0.02	480	18.72± 0.12	18.31± 0.12	17.84± 0.15
155.91	460	20.07± 0.06	19.62± 0.02	19.39± 0.03	19.13± 0.02	480	18.79± 0.08	18.31± 0.11	17.87± 0.15
160.33	460	20.05± 0.04	19.65± 0.02	19.36± 0.02	19.15± 0.02	480	18.75± 0.10	18.35± 0.10	18.01± 0.16
164.70	460	20.14± 0.05	19.75± 0.04	19.43± 0.03	19.19± 0.04	480	18.75± 0.11	18.40± 0.12	18.09± 0.18
239.81	919	20.81± 0.04	20.35± 0.03	20.11± 0.02	19.89± 0.05	960	19.66± 0.13	19.01± 0.14	18.71± 0.17
250.95	919	20.85± 0.06	20.49± 0.02	20.20± 0.03	19.95± 0.07	960	19.65± 0.11	19.11± 0.12	18.98± 0.21
329.17	1133	21.16± 0.06	20.74± 0.03	20.43± 0.03	20.22± 0.05	1920	19.87± 0.08	19.39± 0.12	19.10± 0.18
415.47	1838	21.49± 0.05	21.08± 0.03	20.81± 0.02	20.60± 0.04	1920	20.17± 0.11	19.88± 0.16	19.65± 0.26
501.08	1838	21.59± 0.03	21.19± 0.02	20.90± 0.02	20.71± 0.04	1920	20.25± 0.11	19.99± 0.15	19.94± 0.32
588.10	1838	21.85± 0.05	21.46± 0.03	21.18± 0.04	20.94± 0.05	1920	20.62± 0.16	20.25± 0.19	19.67± 0.27
669.18	919	22.03± 0.05	21.67± 0.08	21.40± 0.09	21.17± 0.08	960	20.70± 0.23	20.36± 0.29	19.86± 0.35
843.66	1379	22.39± 0.03	22.01± 0.03	21.75± 0.04	21.57± 0.06	1440	21.23± 0.21	20.71± 0.40	20.49± 0.46
1101.93	2420	22.86± 0.06	22.42± 0.04	22.20± 0.07	22.03± 0.09	2160	21.83± 0.24	20.82± 0.25	20.57± 0.52
1880.55	2952	23.26± 0.09	22.68± 0.05	22.40± 0.07	22.30± 0.09	2400	21.79± 0.24	21.76± 0.27	20.70± 0.75
2401.32	4502	23.45± 0.19	23.00± 0.09	22.63± 0.11	22.36± 0.14	3600	22.15± 0.32	21.86± 0.36	>20.32
3090.97	3630	23.80± 0.12	23.11± 0.08	22.81± 0.10	22.46± 0.12	3240	>22.25	>21.85	>20.22
4258.44	5384	24.27± 0.24	23.60± 0.13	23.26± 0.23	23.00± 0.32	4560	>21.54	>21.05	>19.19
4732.20	5422	24.47± 0.35	23.92± 0.15	23.38± 0.16	23.15± 0.21	4560	>22.06	>21.62	>20.33
6241.88	2758	>24.57	24.45± 0.28	23.68± 0.32	23.47± 0.44	2880	>21.52	>20.91	>20.06
24277.46	3752	25.66± 0.31	25.04± 0.18	24.36± 0.22	24.02± 0.28	3600	>22.39	>21.84	>20.56

The Δt time gives the mid-time of the observation relative to the Swift trigger time, and T_{VIS} and T_{NIR} are the exposure times in the g′r′I′z and JHK$_s$ filters, respectively. All magnitudes are in the AB system and not corrected for Galactic foreground extinction. Conversion to Vega magnitudes: g′$_{AB}$ − g′$_{Vega}$ = −0.062 mag, r′$_{AB}$ − r′$_{Vega}$ = 0.178 mag, i′$_{AB}$ − i′$_{Vega}$ = 0.410 mag, i′$_{AB}$ − i′$_{Vega}$ = 0.543 mag, J$_{AB}$ − J$_{Vega}$ = 0.929 mag, H$_{AB}$ − H$_{Vega}$ = 1.394 mag, K$_{S,AB}$ − K$_{S,Vega}$ = 1.859 mag. Corrections for Galactic extinction are $A_{g'}$ = 0.066 mag, $A_{r'}$ = 0.046 mag, $A_{i'}$ = 0.034 mag, $A_{z'}$ = 0.025 mag, A_J = 0.015 mag, A_H = 0.010 mag, A_{K_s} = 0.006 mag.

Extended Data Table 2 | UVOT observations of the afterglow of GRB 111209A

Δt (ks)	T (s)	u (mag)
139.3566	546.0	$20.23^{+0.11}_{-0.10}$
187.4401	157.0	$21.14^{+0.77}_{-0.45}$
199.3795	157.0	$21.24^{+0.58}_{-0.38}$
211.8172	157.0	$21.72^{+0.77}_{-0.45}$
223.9091	235.5	$21.25^{+0.42}_{-0.30}$
233.6637	235.5	$21.75^{+0.90}_{-0.49}$
245.1895	156.9	$20.82^{+0.55}_{-0.36}$
256.7393	157.0	$21.74^{+2.17}_{-0.68}$
286.4793	84.7	>20.66
315.6230	314.1	$21.84^{+0.70}_{-0.42}$
332.6649	382.4	$21.98^{+0.52}_{-0.35}$
357.8214	844.0	$21.78^{+0.51}_{-0.34}$
428.4023	578.3	$22.05^{+0.44}_{-0.31}$
465.3887	342.0	$21.45^{+0.42}_{-0.30}$

The Δt time gives the mid-time of the observation relative to the Swift trigger time, and all magnitudes are in the AB system and not corrected for Galactic foreground extinction. Conversion to Vega magnitudes: $u_{AB} - u_{Vega} = 1.02$ mag (as given at http://swift.gsfc.nasa.gov/analysis/uvot_digest/zeropts.html). The correction for Galactic extinction, using $E_{(B-V)} = 0.017$ mag[36] and the Galactic extinction curve[46] is $A_u = 0.085$ mag.

LETTER

Global–scale coherence modulation of radiation–belt electron loss from plasmaspheric hiss

A. W. Breneman[1], A. Halford[2], R. Millan[2], M. McCarthy[3], J. Fennell[4], J. Sample[5], L. Woodger[2], G. Hospodarsky[6], J. R. Wygant[1], C. A. Cattell[1], J. Goldstein[7], D. Malaspina[8] & C. A. Kletzing[6]

Over 40 years ago it was suggested that electron loss in the region of the radiation belts that overlaps with the region of high plasma density called the plasmasphere, within four to five Earth radii[1,2], arises largely from interaction with an electromagnetic plasma wave called plasmaspheric hiss[3–5]. This interaction strongly influences the evolution of the radiation belts during a geomagnetic storm, and over the course of many hours to days helps to return the radiation-belt structure to its 'quiet' pre-storm configuration. Observations have shown that the long-term electron-loss rate is consistent with this theory but the temporal and spatial dynamics of the loss process remain to be directly verified. Here we report simultaneous measurements of structured radiation-belt electron losses and the hiss phenomenon that causes the losses. Losses were observed in the form of bremsstrahlung X-rays generated by hiss-scattered electrons colliding with the Earth's atmosphere after removal from the radiation belts. Our results show that changes of up to an order of magnitude in the dynamics of electron loss arising from hiss occur on timescales as short as one to twenty minutes, in association with modulations in plasma density and magnetic field. Furthermore, these loss dynamics are coherent with hiss dynamics on spatial scales comparable to the size of the plasmasphere. This nearly global-scale coherence was not predicted and may affect the short-term evolution of the radiation belts during active times.

We analyse in detail the magnetic conjunctions between the Van Allen probes (a NASA Earth-orbiting satellite mission)[6] and the Balloon Array for Radiation belt Relativistic Electron Losses (BARREL) balloons[7] flying at altitudes of nearly 35 km over Antarctica on 3 and 6 January 2014. We then used this new set of measurements to explore spatial and temporal scales previously not easily accessible. On 2 January 2014, a dynamic injection of energetic electrons (tens of kiloelectronvolts; thought to arise from lower-energy populations beyond the radiation belts) into the outer radiation belt down to at least $L = 3$ occurred. (L is defined as a magnetic field line whose value is the distance in Earth radii (R_E) at which it intersects the magnetic equator.) These electrons, trapped within the plasmasphere, were gradually eroded over the course of a few days by various loss mechanisms, including interaction with plasmaspheric hiss. The hiss on 3 and 6 January has a peak power at frequencies <100 Hz and probably originates from a free-energy source within the plasmasphere[8–10].

Figure 1 plots three hours of detrended hiss and X-ray observations, corresponding to electrons with energies of tens of kiloelectronvolts for an afternoon-sector conjunction between Van Allen probe P_A and balloon B_I on 3 January 2014. For a detailed description of a conjunction see the Methods and Extended Data Fig. 1. During this conjunction B_I stays near the field line $L = 5.2$, which maps to the plasmasphere, while P_A traverses a large extent of the afternoon-sector plasmasphere from $L = 5.5$–2.5. It is immediately evident that there is

a strong visible coherence between the hiss amplitude and the X-ray count rate at periods of 1–20 min throughout the conjunction. This coherence extends throughout the entire afternoon-sector plasmasphere, even when P_A and B_I have spatial separations of up to $3L$ and nearly 3 h of magnetic local time (MLT), corresponding to separations across the magnetic field in the range $2.3R_E$–$4.1R_E$, or 14,500–26,000 km.

Similar long-distance coherence was observed for a few days after the energetic particle injection on 2 January. Six payloads (P_A, P_B, B_K,
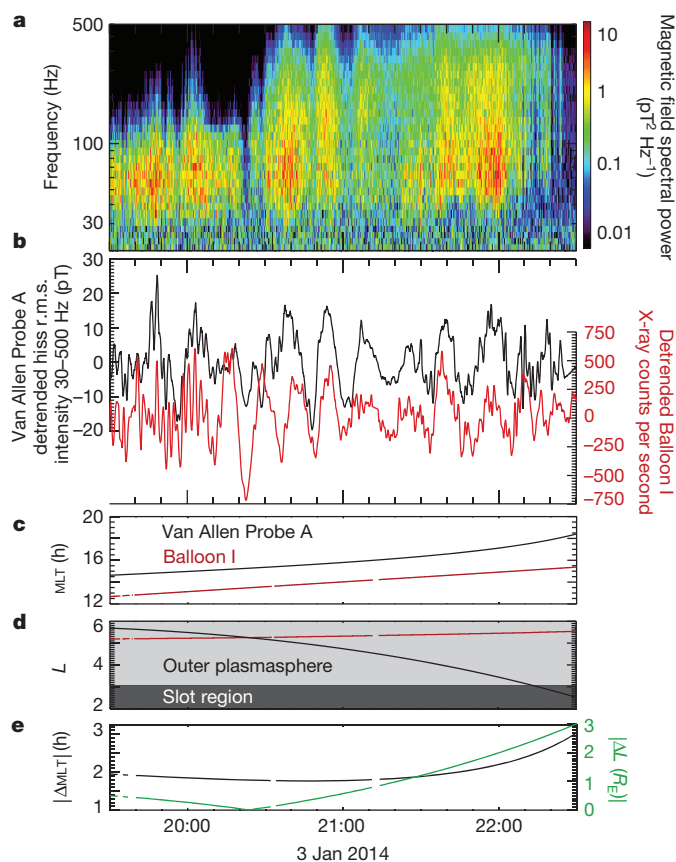
Figure 1 | Comparison of satellite and balloon data showing large-scale coherence. A magnetic field spectrogram (**a**) from the Electric Fields and Waves instrument[10] on P_A shows hiss with peak power at frequencies <100 Hz (ref. 9). The root mean square (r.m.s.) hiss amplitude, up to 20 pT and consistent with quiet-time values[8], shows fluctuations similar to the X-ray count rate on B_I (**b**), caused by electrons of 10–200 keV. Both curves were detrended with a 20-min running boxcar average to emphasize fluctuations. Panels **c**–**e** show the MLT, L, $|\Delta$MLT$|$ and $|\Delta L|$ values for P_A and B_I.

[1]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA. [2]Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA. [3]Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA. [4]Aerospace Corporation, Los Angeles, California 90009, USA. [5]Space Sciences Laboratory, University of California, Berkeley, California 94720, USA. [6]Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA. [7]Southwest Research Institute, San Antonio, Texas 78238, USA. [8]University of Colorado, The Laboratory for Atmospheric and Space Physics (LASP), Boulder, Colorado 80303, USA.
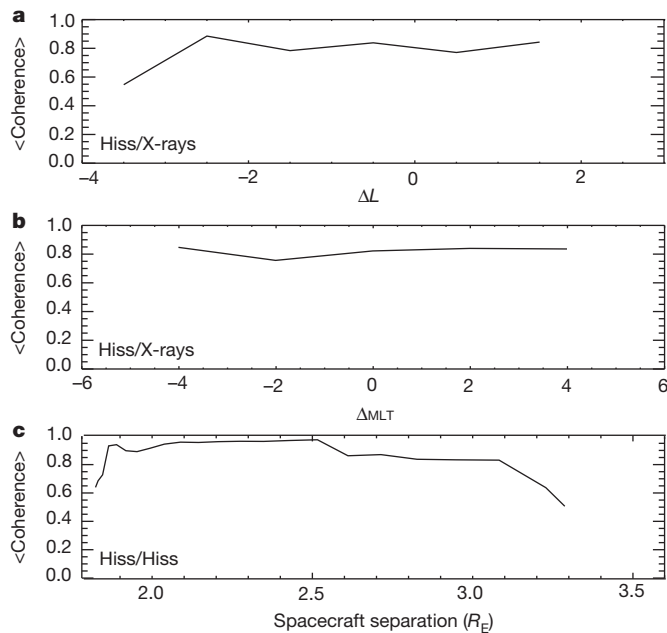
Figure 2 | The extent over which the hiss source region is modulated. Time-dependent coherence of hiss and X-rays on 6 January 2014 at 20:00–22:00 UT was calculated as a function of cross-magnetic-field separation for a range of fluctuation periods. High coherence is observed for 1–20-min periods. This figure shows a coherence of 3.3-min period fluctuations averaged for all combinations of P_A and P_B and B_K, B_L, B_W as a function of L (**a**) and MLT (**b**). Panel **c** plots the coherence of 3.3-min period fluctuations of hiss as a function of spacecraft separation.

Figure 3 | Comparison of plasmaspheric hiss and electron precipitation. **a**, A spectrogram of hiss magnetic field wave power along the spin axis of P_A. The r.m.s. hiss power and X-ray count rate on B_K follow the same trend (**b**). The background X-ray count rate, probably due to cosmic-ray-induced X-rays and estimated from a 20-min running boxcar average to be 875 counts per second, has been subtracted. Plasma density, 30-keV electron flux, and magnetic field magnitude B (**c**–**e**) all increase during this conjunction, unlike for the X-ray counts.

B_W, B_L, B_X) were aloft during the 6 January 2014 conjunction and on magnetic field lines mapping to the plasmasphere. This data set allows us to quantify the spatial extent over which fluctuations of hiss and X-rays are similar. Figure 2a and b plots the average coherence (strength of association) of 3.3-min period fluctuations of hiss amplitude and X-ray count rate as a function of cross-magnetic-field separation for all combinations of a probe and a balloon. Between any two payloads, high coherence was observed for separations up to $|\Delta\text{MLT}| = 4$ h and $|\Delta L| = 3.5$. However, since high coherence was observed on all combinations of payloads, the overall coherence scale covers, at a minimum, all baselines formed by the probes and balloons: that is, 6 h of MLT (from $11:00 < \text{MLT} < 17:00$) and $3.5L$ (from $3 < L < 6.5$). A comparison in Fig. 2c of hiss amplitudes on P_A and P_B shows that 3.3-min period fluctuations of the hiss source region are coherent on similar scales, indicating that the large-scale coherence of hiss and X-rays may be explained by the large-scale coherence of the hiss source region itself. We note that the large-scale MLT coherence is not caused by magnetic-field gradient and curvature drift of electrons into the South Atlantic Anomaly[11], a region of decreased magnetic field associated with enhanced precipitation from the radiation belts. Timescales for the drift of 30–100-keV electrons with small pitch angles, the angle of the electron velocity vector to the magnetic field, range from a few to tens of minutes per hour of MLT. Delays in fluctuations in hiss amplitude and X-ray counts of this size are not observed.

Figure 3 presents a detailed analysis of this 6 January close conjunction for the payloads P_A and B_K. As for the 3 January conjunction, hiss amplitude and X-ray count rate show a striking similarity throughout the entire two-hour conjunction. The payload separations are small near 21:00 Universal Time (UT) ($\Delta\text{MLT} = 1$ h and $\Delta L = 1.5$), meaning that these variations are mostly temporal. The plasma density and magnetic-field magnitude, as well as the 30-keV electron flux, rise on average throughout this timespan, unlike the X-ray count rate, although when detrended they fluctuate in the ultralow frequency (ULF) range, similarly to the hiss and X-rays. In a detailed analysis

presented in the Methods we provide multiple lines of evidence showing that the hiss is directly responsible for creating the electron precipitation that enhances the X-rays. The global-scale coherence observed during this conjunction indicates that hiss controls electron loss throughout the afternoon-sector plasmasphere. The ULF fluctuations in density and magnetic field facilitate this process by creating conditions favourable for the growth of hiss waves. These results explain fluctuations in X-ray counts that have been observed on past balloon missions[12].

The ULF fluctuations of density and magnetic field may originate in the solar wind or at the magnetopause boundary[13], within the low-plasma-density magnetosphere[14], or within the plasmasphere[15]. We have not identified their source for the 3 January conjunction but the distinctive double-peaked feature on 6 January at 21:00 UT (Fig. 3b) is observed as a small spike in the SYM-H index, suggesting a compression at the magnetopause, and on the Cluster 4 satellite[16] located in the afternoon-sector magnetosheath. Multiple ground magnetometer stations[17] show that this ULF feature propagates throughout the afternoon-sector plasmasphere, with velocity components eastwards and radially inwards, perturbing the background density and magnetic field by a few per cent. This has a noticeable effect on the growth of plasmaspheric hiss, causing an increase in electron precipitation of up to an order of magnitude observed throughout the afternoon-sector plasmasphere all the way inward to the outer edge of the electron slot at $L = 3$, a region largely devoid of energetic electrons separating the inner and outer radiation belts.

These observations suggest that coupling models of ULF wave formation and propagation to current radiation-belt models is

important to allow accurate simulation of electron loss caused by plasmaspheric hiss during active times. Furthermore, recent results[18] have shown that plasmaspheric hiss waves are occasionally coherent and may have dispersive frequency signatures. Many models assume featureless, broadband hiss, which does not accurately describe pitch-angle scattering of electrons due to coherent, dispersive hiss waves. Coherent hiss waveforms are most prevalent during times when the plasmasphere is driven by solar wind or magnetosheath ULF pressure fluctuations[19], similar to what produces the large-scale coherence for the 6 January conjunction.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Gringauz, K. I. The structure of the ionized gas envelope of Earth from direct measurements in the USSR of local charged particle concentrations. *Planet. Space Sci.* **11,** 281–296 (1963).
2. Carpenter, D. L. Whistler evidence of a "knee" in the magnetospheric ionization density profile. *J. Geophys. Res.* **68,** 1675–1682 (1963).
3. Lyons, R. L., Thorne, R. M. & Kennel, C. F. Pitch-angle diffusion of radiation belt electrons within the plasmasphere. *J. Geophys. Res.* **77,** 3455–3474 (1972).
4. Lyons, L. & Thorne, R. M. Equilibrium structure of radiation belt electrons. *J. Geophys. Res.* **78,** 2142–2149 (1973).
5. Millan, R. M. & Thorne, R. M. Review of radiation belt relativistic electron losses. *J. Atmos. Sol. Terr. Phys.* **69,** 362–377 (2007).
6. Mauk, B. H. *et al.* Science objectives and rationale for the Radiation Belt Storm Probes Mission. *Space Sci. Rev.* **179,** 3–27 (2013).
7. Millan, R. M. *et al.* The Balloon Array for RBSP Relativistic Electron Losses (BARREL). *Space Sci. Rev.* **179,** 503–530 (2013).
8. Li, W. *et al.* An unusual enhancement of low-frequency plasmaspheric hiss in the outer plasmasphere associated with substorm-injected electrons. *Geophys. Res. Lett.* **40,** 3798–3803 (2013).
9. Chen, L. *et al.* Generation of unusually low frequency plasmaspheric hiss. *Geophys. Res. Lett.* **41,** 5702–5709 (2014).
10. Ni, B. *et al.* Resonant scattering of energetic electrons by unusual low-frequency hiss. *Geophys. Res. Lett.* **41,** 1854–1861 (2014).
11. Pinto, O. *et al.* The South Atlantic Magnetic Anomaly: three decades of research. *J. Atmos. Sol. Terr. Phys.* **54,** 1129–1134 (1992).
12. Millan, R. M. *et al.* Understanding relativistic electron losses with BARREL. *J. Atmos. Sol. Terr. Phys.* **73,** 1425–1434 (2011).
13. Claudepierre, S. G., Hudson, M. K., Lotko, W., Lyon, J. G. & Denton, R. E. Solar wind driving of magnetospheric ULF waves: field line resonances driven by dynamic pressure fluctuations. *J. Geophys. Res.* **115,** A11202 (2010).
14. Hughes, W. J., Southwood, D. J., Mauk, B., McPherron, R. L. & Barfield, J. N. Alfven waves generated by an inverted plasma energy distribution. *Nature* **275,** 43–45 (1978).
15. Dai, L. *et al.* Excitation of poloidal standing Alfven waves through drift resonance wave-particle interaction. *Geophys. Res. Lett.* **40,** 4127–4132 (2013).
16. Escoubet, C. P., Schmidt, R. & Goldstein, M. L. Cluster—science and mission overview. *Space Sci. Rev.* **79,** 11–32 (1997).
17. Mann, I. R. *et al.* The upgraded CARISMA magnetometer array in the THEMIS era. *Space Sci. Rev.* **141,** 413–451 (2008).
18. Summers, D., Omura, Y., Nakamura, S. & Kletzing, C. A. Fine structure of plasmaspheric hiss. *J. Geophys. Res.* **119,** 9134–9149 (2014).
19. Tsurutani, B. T., Falkowski, B. J., Pickett, J. S., Santolik, O. & Lakhina, G. S. Plasmaspheric hiss properties: observations from Polar. *J. Geophys. Res.* **120,** 414–431 (2015).

METHODS

In the absence of external electric fields, magnetospheric electrons have helical trajectories, which are the result of gyration about the magnetic field and translation of the centre of this gyration—the guiding centre—along the magnetic field. The cone of velocity vectors representing these two motions is inclined from the magnetic field by an angular distance called the pitch angle. As electrons propagate along magnetic field lines towards higher latitudes, they encounter a mirror force, caused by the convergence of magnetic field lines, that opposes their guiding-centre velocity. For electrons with pitch angles outside what is called the bounce loss cone, this force is sufficient to reverse their guiding-centre motion before they reach the atmosphere. These electrons are trapped locally within the magnetosphere, bouncing between mirror points in opposite hemispheres. Electrons with pitch angles inside the bounce loss cone reach the atmosphere before mirroring and are lost from the radiation belts.

Extended Data Fig. 1 illustrates a typical conjunction between a BARREL balloon and the Van Allen probes P_A and P_B. P_A resides on a magnetic field line inside of the balloon field of view. Electrons scattered into the bounce loss cone by hiss waves within this field of view enter the atmosphere, where they create bremsstrahlung X-rays visible to the conjugate balloon. P_B resides outside of the field of view and bounce-loss-cone electrons at this location will not be observed on the balloon unless some intermediate process modulates electron loss over long distances across the magnetic field. We now discuss the mechanism by which this scattering occurs and why ULF-period enhancements of plasma density and decreases in magnetic field provide this modulation and enhance the loss rate.

Plasmaspheric hiss interacts with electrons via a Doppler-shifted cyclotron resonance. This interaction can scatter trapped electrons into the bounce loss cone, removing them from the radiation belts. Strong resonance occurs when an electron, propagating with a guiding centre velocity V towards a hiss wave with frequency f in the plasma frame, defined in terms of the angular frequency as $f = \omega/2\pi$ and wavenumber k, observes the hiss wave Doppler-shifted to an integral multiple n of its gyration frequency, that is, the cyclotron frequency $f_{ce} = \omega_{ce}/2\pi$. The resonance condition is given by $n\omega_{ce} = \omega - \delta\omega$, where $\delta\omega = kV/\gamma$ is the Doppler shift and γ is the relativistic mass factor[20]. An electron in resonance is exposed to a coherent hiss magnetic and electric field and experiences a net exchange of momentum and energy with the wave. This is possible because, in a frame that is moving together with the electron-guiding centre, the counter-streaming electron and hiss wave have the same sense of rotation about the magnetic field. For interaction with typical hiss waves, the electron energy remains constant and so the change in momentum results in a change in electron pitch angle.

Since the electron guiding centre velocity is typically along the magnetic field direction, $kV = kV\cos\theta$, where θ is the angle the wavenumber vector makes to the magnetic field, called the wave normal angle. Wave normal angles are determined on the Van Allen probes from a singular value decomposition analysis[21] supplied by the Electric and Magnetic Field Instrument Suite and Integrated Science instrument (EMFISIS)[22]. The hiss observed during the 3 and 6 January conjunctions is relatively field-aligned throughout the hiss band, with θ in the range 20°–40°. This indicates that first-order Doppler-shifted cyclotron resonance ($n = 1$) dominates higher-order resonances.

Using *in situ* values of density, magnetic field, hiss frequency and θ we determine that first-order cyclotron resonance energies range from a few tens of kiloelectronvolts up to a few hundred kiloelectronvolts for both conjunctions. Extended Data Fig. 2a shows these energies plotted as a function of L (determined from the T89 magnetic field model[23] with an input planetary index of $K_P = 1$) for P_A on 6 January. Resonance energies tend to dip during times of enhanced hiss power, primarily due to 1–20-min fluctuations in density (enhancements) and magnetic field (depressions) which are often observed at the same time. Extended Data Fig. 2b shows Magnetic Electron Ion Spectrometer (MagEIS) instrument[24] observations on P_A of differential electron flux in a range of energy channels, plotted as a function of L. Under the assumption that these fluxes do not substantially change over the duration of the conjunction, we take these values to represent the fluxes in the hiss source region at the L value of each balloon. The range of electron energies required to satisfy the local first-order cyclotron resonance condition at the location of each balloon is indicated by the shaded rectangles. B_L and B_K, which observe substantial X-ray counts, map to an area of the hiss source region where MagEIS observed enhanced electron populations at energies required for resonance. B_X, which sees only background level X-ray counts, maps to a lower L where MagEIS observed few electrons at the required higher resonance energies. A similar result is obtained for the 3 January conjunction where B_I observed substantial electron precipitation but B_W does not. Hiss scattering of electrons into the loss cone via the mechanism of Doppler-shifted first-order cyclotron resonance is therefore consistent with the balloon observations.

Each individual resonance interaction is equally likely to scatter an electron towards higher or lower pitch angle. Because electron pitch-angle distributions in the plasmasphere are typically anisotropic, with more electrons at higher than lower pitch angles, the net angular scattering over time skews towards lower pitch angles in a manner thought to be consistent with a diffusive process. The simultaneous satellite and balloon observations allow comparison of the theoretical loss rates to observed loss rates on the balloons. We now show that these two rates are compatible for the close conjunction on 6 January near 21:00 UT.

The pitch angle α through which an electron will random walk in a single bounce period T_b due to interaction with hiss is given by $\Delta\alpha = \sqrt{2\langle D_{\alpha\alpha}\rangle T_b}$, where $\langle D_{\alpha\alpha}\rangle$ is the bounce-averaged pitch-angle diffusion rate, calculated from *in situ* density, magnetic field magnitude, θ, amplitude and bandwidth[25]. We have approximated the hiss frequency spectrum with a Gaussian, centred on the frequency of peak amplitude at each time with a bandwidth determined by the minimum and maximum extent of the observed hiss wave power. The observed spectrum has an extended high-frequency tail not matched by a normal Gaussian, but for the purposes of showing the consistency between the predicted flux at P_A and the observed flux on B_K a Gaussian is sufficient. Modelled hiss waves are allowed to interact with electrons with energies from 0–200 keV and with $\alpha = 0°-30°$ in a static dipole magnetic field. Diffusion rates, averaged over the gyration period of the electron about the magnetic field $D_{\alpha\alpha}$, were determined at each point along the bounce path. Next, bounce-averaged diffusion rates $\langle D_{\alpha\alpha}\rangle$ were calculated as a function of energy and time by averaging the diffusion rates for electrons with pitch angles near the bounce loss cone over the entire electron bounce path. Because $\langle D_{\alpha\alpha}\rangle$ is a function of not only hiss magnetic field wave power, but also of density and background magnetic field, it is modified by the ULF-period fluctuations.

The BARREL field of view mapped to the magnetic equator is a circle of approximate diameter $1 R_E$, and X-rays observed at B_K will be produced by precipitating electrons within this entire field of view. Exactly determining the overall loss rate would require knowledge of the instantaneous hiss amplitude distribution over this entire field of view, which is clearly not possible without multiple satellite measurements. We can, however, approximate the effect of such a distribution by averaging $\langle D_{\alpha\alpha}\rangle$ over the 15 min it takes P_A to cross this field of view, resulting in a scattering rate for 50-keV electrons of 10^{-4} s^{-1}. Using this time-averaged value, 50-keV electrons within the field of view of B_K will scatter an angular distance of approximately 1° over a typical bounce period of 1 s.

From MagEIS data we can estimate the number of electrons within 1° of the 4° bounce loss cone. During the conjunction near 20:58 UT the MagEIS instrument is able to make high-time-resolution measurements to within 2.5° of the bounce loss cone. Considering the time measuring each sector is small there is large error associated with these values. Good counting statistics are obtained by averaging over ten spin periods. Results indicate a range of 25–700 electrons (cm^2 s sr keV)$^{-1}$ of energy 50 keV at the bounce loss cone for a 15-min window centred on 20:58 UT. Integrating over the solid angle of electrons within 1° of the bounce loss cone gives us the number of electrons that can theoretically be scattered by the observed hiss into the loss cone in a single bounce period. Roughly half of these electrons, corresponding to 0.08–2.4 electrons (cm^2 s keV)$^{-1}$, will be scattered into the bounce loss cone, and the other half will be scattered towards higher pitch angles. This is the predicted loss rate in the hiss source region due to quasi-linear diffusion.

As these electrons propagate down magnetic field lines from near the magnetic equator to the atmosphere they are focused into a smaller cross-sectional area by the converging magnetic field lines, effectively increasing the differential flux by the magnetic focusing factor $A_{eq}/A_{70 km} = B_{70 km}/B_{eq}$, where A is the cross-sectional area of the field of view of the balloon observing the precipitation and B is the magnetic field strength. Estimates of this focusing factor range from 250 to 310 depending on the variation in field strength over the $1 R_E$-diameter field of view at the magnetic equator. Thus the flux of 0.08–2.4 electrons (cm^2 s keV)$^{-1}$ at 50 keV scattered into the loss cone each bounce will correspond to 15–450 electrons (cm^2 s keV)$^{-1}$ at 70 km, the altitude where the bremsstrahlung X-rays are created. If hiss is indeed the cause of the electron loss then the flux extracted from balloon X-ray counts should be consistent with this number.

To compare this range of values with observations at B_K we invert the X-ray counts at 20:58 UT with a bremsstrahlung X-ray model. A necessary step in this process is the subtraction of the background level of X-ray counts. These can be created from at least two sources, including cosmic rays and enhanced South Atlantic Anomaly precipitation. After background subtraction we find that BARREL observed roughly 13 electrons (cm^2 s keV)$^{-1}$ at 50 keV at 20:58 UT. This is probably an underestimate of the true flux by a factor of approximately three, caused by the assumption by the bremsstrahlung model that precipitating electrons at an altitude of 70 km are isotropically distributed in pitch angle. This will only be the case when the limit of strong diffusion is reached, defined as where the hiss waves are able to fill the bounce loss cone completely in a quarter

bounce period. The diffusion rates calculated are much lower than this and thus only the edge of the bounce loss cone will be filled with electrons. Compensating for this factor, our estimate of 50-keV fluxes observed on B_K is in the range $26–39\,(\mathrm{cm}^2\,\mathrm{s}\,\mathrm{keV})^{-1}$. This is consistent with the predicted flux of 15–450 electrons $(\mathrm{cm}^2\,\mathrm{s}\,\mathrm{keV})^{-1}$. Thus the electron loss associated with the X-ray count rate on B_K during the close conjunction on 6 January near 21 UT can be explained by quasi-linear scattering of electrons into the loss cone by the observed hiss waves.

We end with a power spectral analysis that provides additional evidence that plasmaspheric hiss, modulated by ULF-period fluctuations in density and magnetic field, is directly responsible for precipitating the electrons that create the X-ray fluctuations. Results are presented for the 3 January conjunction because on 6 January strong hiss and X-ray peaks near 21:00 UT can dominate the spectral comparison. However, results are similar for both days. The left column in Extended Data Fig. 3 plots detrended curves of the quantities $\langle D_{\alpha\alpha}\rangle$, hiss amplitude, density, magnetic field and 54-keV electron flux, compared to detrended X-ray counts. Not only are fluctuations in X-ray counts best matched by hiss amplitude and $\langle D_{\alpha\alpha}\rangle$, but so are the spectra shown in the right column. The density and X-ray spectra compare favourably but to a lesser degree, and the magnetic field and energetic electron flux spectra are dissimilar in that they are enhanced at frequencies where the X-ray spectrum is depressed. Taken together, these comparisons indicate that the time and spectral characteristics of the hiss more closely match those of the X-ray counts than do density, magnetic field magnitude and energetic electron flux, strongly suggesting that the hiss is directly responsible for the electron loss. Despite this, enhancements in hiss amplitude tend to occur in regions of enhanced density and depressed magnetic field, which creates conditions favourable for hiss growth[26].

20. Kennel, C. F. & Petschek, H. E. Limit on stably trapped particle fluxes. *J. Geophys. Res.* **71,** 1–28 (1966).
21. Santolík, O., Parrot, M. & Lefeuvre, F. Singular value decomposition methods for wave propagation analysis. *Radio Sci.* **38,** 1010 (2003).
22. Kletzing, C. A. *et al.* The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) on RBSP. *Space Sci. Rev.* **179,** 127–181 (2013).
23. Tsyganenko, N. A. A magnetospheric magnetic field model with a warped tail current sheet. *Planet. Space Sci.* **37,** 5–20 (1989).
24. Blake, J. B. *et al.* The Magnetic Electron Ion Spectrometer (MagEIS) instruments aboard the Radiation Belt Storm Probes (RBSP) spacecraft. *Space Sci. Rev.* **179,** 383–421 (2013).
25. Summers, D., Ni, B. & Meredith, N. P. Timescales for radiation belt electron acceleration and loss due to resonant wave-particle interactions: 1. Theory. *J. Geophys. Res.* **112,** A04207 (2007).
26. Li, W. *et al.* An unusual enhancement of low-frequency plasmaspheric hiss in the outer plasmasphere associated with substorm-injected electrons. *Geophys. Res. Lett.* **40,** 3798–3803 (2013).
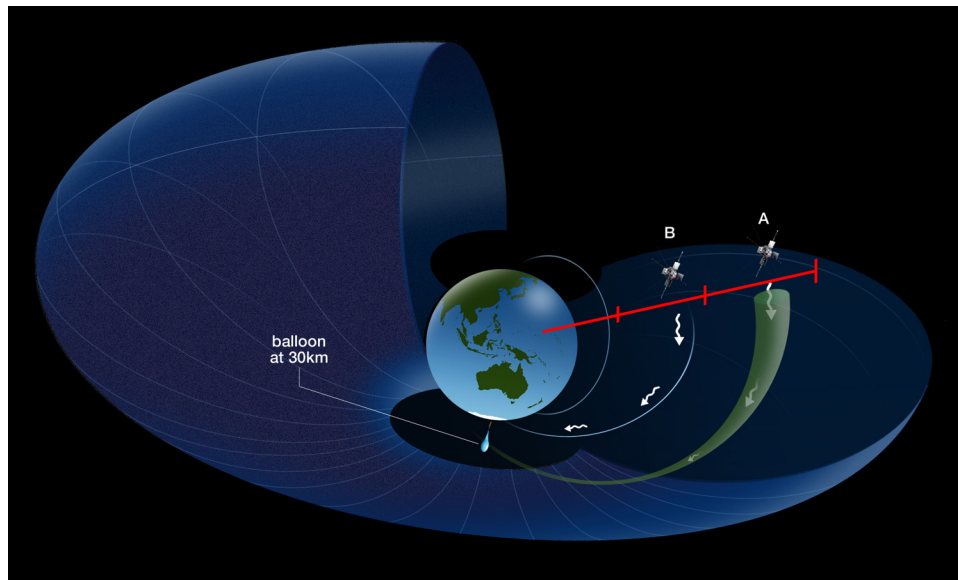
Extended Data Figure 1 | The detection of loss cone electrons by a balloon. The Van Allen probes (labelled here A and B) pass through the hiss source region at the magnetic equator ($3R_E$–$6R_E$, 40,000 km altitude) on field lines that can connect to the BARREL balloons. The red hatched line shows values of $2R_E$, $4R_E$ and $6R_E$. The shaded green volume shows the balloon field of view and the white lines represent electrons propagating along and gyrating about magnetic field lines. At an altitude of 70 km, where bremsstrahlung X-rays are typically created, the cross-section of this field of view is a circle about 100 km in radius. Mapped along magnetic field lines to the magnetic equator, this becomes a circle of radius $0.5R_E$ (about 3,200 km).
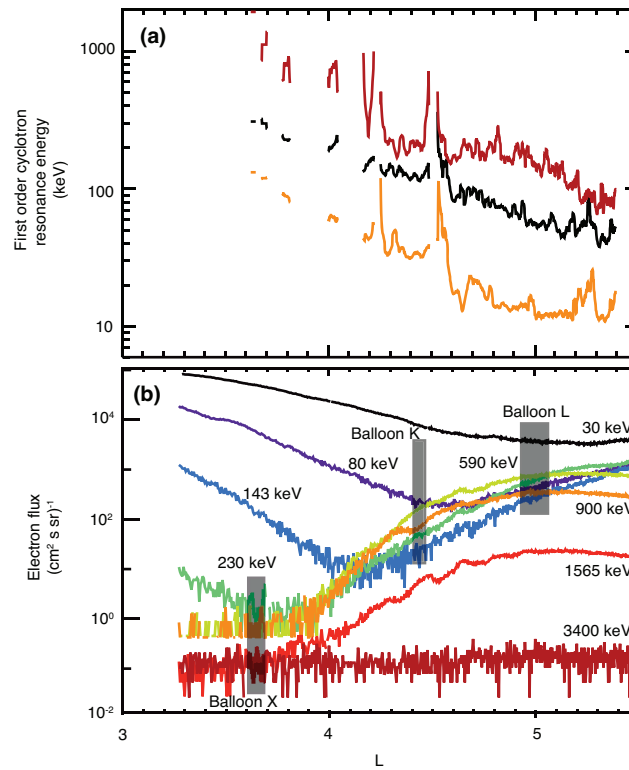
Extended Data Figure 2 | Comparison of predicted resonance energies to observed energies. a, Calculation of first-order cyclotron resonance energies (6 January 2014, 20:00–22:00 UT) from *in situ* data on P_A versus *L*. The three lines are resonant energies determined from the minimum (red), peak (black) and maximum hiss (orange) frequencies. **b,** MagEIS electron flux levels on P_A versus *L*. The horizontal extent of each shaded box shows the *L* crossed by the balloons. The vertical extent is a mapping of the range of resonant energies from **a** across the observed energies from MagEIS.
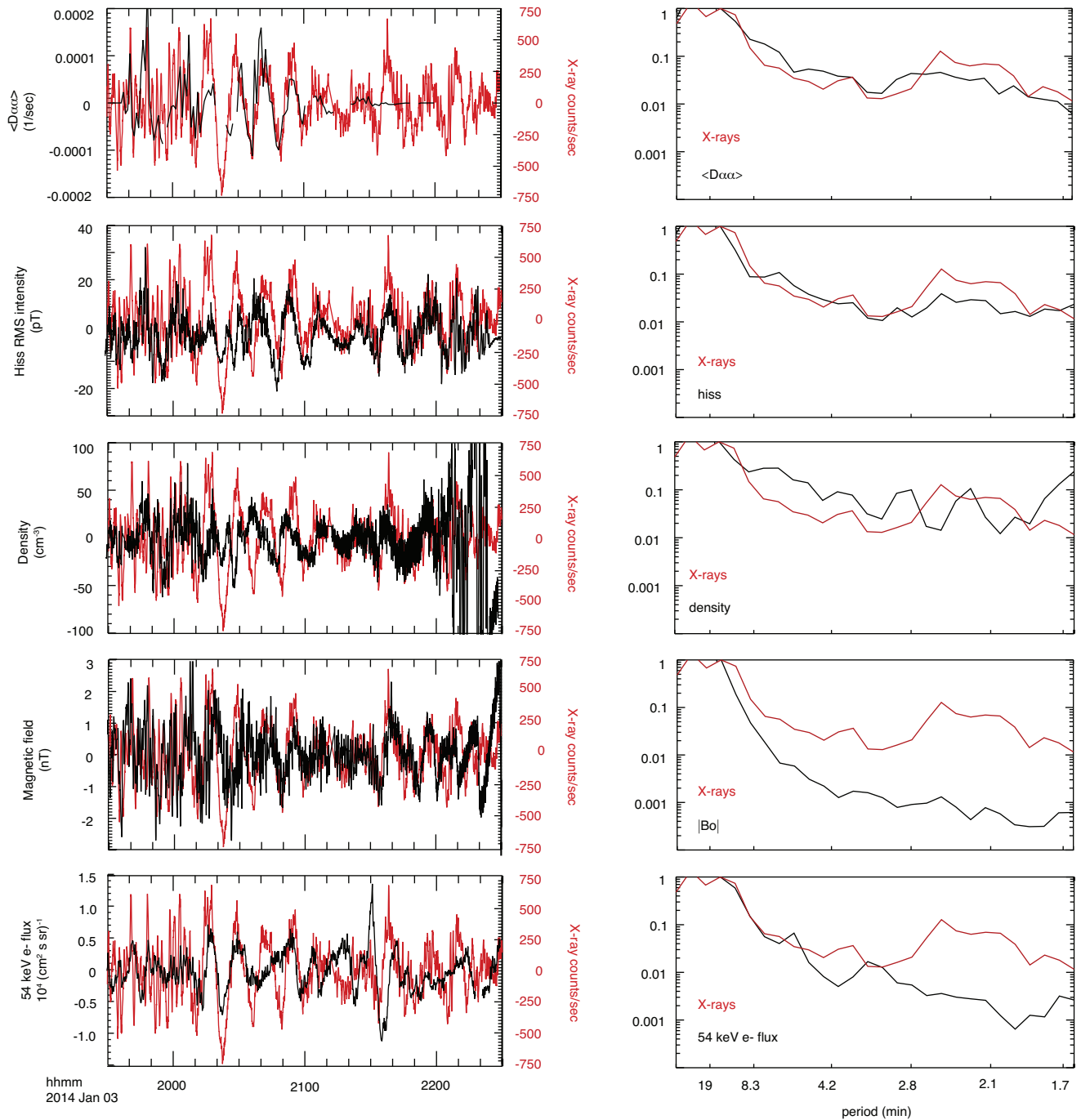
Extended Data Figure 3 | Analysis suggesting that hiss is directly responsible for observed electron loss. Coherence and power spectra from P_A and B_I of fluctuating quantity pairs $\langle D_{\alpha\alpha}\rangle$/X-rays, hiss amplitude/X-rays, magnetic field/X-rays, density/X-rays, and 54-keV electron flux/X-rays. The left plot in each row shows detrended quantity pairs while the right plot is the respective power spectral comparison for fluctuation periods from 1–20 min. To provide comparisons between spectra with different units, values are presented in decibels relative to the power of each curve at the 16.7-min period.

LETTER

Long–range energy transport in single supramolecular nanofibres at room temperature

Andreas T. Haedler[1], Klaus Kreger[1], Abey Issac[2]†, Bernd Wittmann[2], Milan Kivala[3], Natalie Hammer[3], Jürgen Köhler[2], Hans–Werner Schmidt[1] & Richard Hildner[2]

Efficient transport of excitation energy over long distances is a key process in light-harvesting systems, as well as in molecular electronics[1–3]. However, in synthetic disordered organic materials, the exciton diffusion length is typically only around 10 nanometres (refs 4, 5), or about 50 nanometres in exceptional cases[6,7], a distance that is largely determined by the probability laws of incoherent exciton hopping. Only for highly ordered organic systems has the transport of excitation energy over macroscopic distances been reported—for example, for triplet excitons in anthracene single crystals at room temperature[8], as well as along single polydiacetylene chains embedded in their monomer crystalline matrix at cryogenic temperatures (at 10 kelvin, or −263 degrees Celsius)[9]. For supramolecular nanostructures, uniaxial long-range transport has not been demonstrated at room temperature. Here we show that individual self-assembled nanofibres with molecular-scale diameter efficiently transport singlet excitons at ambient conditions over more than four micrometres, a distance that is limited only by the fibre length. Our data suggest that this remarkable long-range transport is predominantly coherent. Such coherent long-range transport is achieved by one-dimensional self-assembly of supramolecular building blocks, based on carbonyl-bridged triarylamines[10], into well defined H-type aggregates (in which individual monomers are aligned cofacially) with substantial electronic interactions. These findings may facilitate the development of organic nanophotonic devices and quantum information technology.

Supramolecular chemistry uses directed noncovalent interactions between molecules to construct well defined architectures that provide functionalities beyond those of the constituent molecular building blocks[11–16]. For example, supramolecular nanostructures have been identified as potential components for transporting excitation energy in light-harvesting applications[17–22] such as solar cells. These nanostructures feature a controlled and spatially well defined arrangement of their building blocks, with substantial intermolecular electronic coupling, which is a requirement for efficient energy transport. The observed energy-migration distances of around 100 nm have been attributed to incoherent exciton hopping with some contribution of coherent motion—that is, a delocalization of electronic excitations over several building blocks. In those nanostructures J-type aggregates are formed, by a brick-layer-type arrangement of the constituent molecules[17–20,22]. In J-aggregates, however, the oscillator strength is redistributed into the lowest-energy exciton level, which forms a super-radiant state[23] with a much shorter excited-state lifetime with respect to that of the noninteracting building blocks. Consequently, the competition between radiative decay and energy transport strongly constrains the distance over which electronic excitations can migrate. In contrast, in ideal H-aggregates, with cofacially stacked building blocks, the lowest-energy transition is dipole forbidden—that is, optically not accessible[23]. This results in a strongly increased radiative lifetime for the relaxed excited state, which

should be beneficial for efficient energy transport over macroscopic distances[8,24].

Here we demonstrate, at room temperature, long-range energy transport along individual H-aggregated nanofibres of molecular-scale diameter. We use a specifically designed supramolecular building block (compound **1**, Fig. 1a) that has a carbonyl-bridged triarylamine (CBT) as its core; this core is functionalized at positions 2, 6 and 10 via amide linkers with 4-(5-hexyl-2,2′-bithiophene)-naphthalimides (NIBTs)[10]. This combination of a planar, aromatic heterotriangulene

Figure 1 | Self-assembly of compound 1. Green, carbonyl-bridged triarylamine (CBT) core; blue, amide moieties; yellow/gold, 4-(5-hexyl-2,2′-bithiophene)-naphthalimide (NIBT) periphery. **a**, Chemical structure of compound **1**. **b**, Self-assembly into nanofibres with an ordered H-aggregated core, driven by π-stacking of CBTs and stabilized by three chains of hydrogen bonds between the amide groups.

[1]Macromolecular Chemistry I, Bayreuth Institute of Macromolecular Research, and Bayreuth Center for Colloids and Interfaces, University of Bayreuth, 95440 Bayreuth, Germany. [2]Experimental Physics IV and Bayreuth Institute of Macromolecular Research, University of Bayreuth, 95440 Bayreuth, Germany. [3]Chair of Organic Chemistry I, Department of Chemistry and Pharmacy, University of Erlangen–Nürnberg, 91054 Erlangen, Germany. †Present address: Department of Physics, Sultan Qaboos University, 123 Muscat, Oman.

core[25] and three hydrogen-bonding amide groups is the structure-defining element that enforces columnar self-assembly (Fig. 1b). The peripheral NIBT units exhibit a bright orange photoluminescence either upon direct photoexcitation or after energy transfer from CBT.

The pronounced self-assembly behaviour of compound **1** is demonstrated by gelation at a concentration of 700 μM (1,000 parts per million, p.p.m.) in *ortho*-dichlorobenzene (*o*-DCB)[10]. We used transmission electron microscopy (TEM) to study a solvent-free sample, directly prepared from the gel, and saw a dense network of nanofibres (Fig. 2a). Some fibres align next to each other, resulting in structures with widths of multiples of 5 nm (Supplementary Fig. 1). The mean diameter of single fibres of 5 ± 1 nm is close to the calculated diameter of 6 nm for compound **1** (Fig. 2b, c), revealing the presence of one-dimensional nanofibres with molecular-scale diameter. This finding is confirmed by the uniform fibre heights of 2–2.5 nm observed with atomic force microscopy (AFM) on samples spin-coated from a dispersion of self-assembled compound **1** in *o*-DCB (7 μM, 10 p.p.m.; Supplementary Fig. 2).

The self-assembly into nanofibres is driven by π-stacking of the aromatic CBT units (that is, by their cofacial arrangement driven by van der Waals interactions between their π-electron systems), as shown by changes in the optical spectra (Supplementary Fig. 3). In the absorption spectra, a strongly blueshifted band arises at about 380 nm (around $26,300\ \mathrm{cm}^{-1}$) upon self-assembly in *o*-DCB, indicating the formation of H-aggregates. However, it is not clear which chromophores are involved in this process, because the absorption spectra of CBT and NIBT overlap[10]. In contrast, the photoluminescence of self-assembled compound **1** stems exclusively from the NIBT chromophores (Supplementary Fig. 3). The photoluminescence spectra do not feature new and strongly shifted bands that can be associated

with aggregation. Hence, the NIBT units are not stacked in an ordered fashion. Only the CBT units within the supramolecular nanofibres form well defined π-stacked assemblies. These one-dimensional H-aggregates are stabilized by intermolecular hydrogen bonds between the amide groups of compound **1**, as shown by changes in the N–H stretch vibration in Fourier-transform infrared spectra (Supplementary Fig. 4). Carbonyl-bridged triarylamines without amide groups are stacked with a clear offset[25,26]. In compound **1**, however, the three chains of hydrogen bonds enforce the cofacial arrangement of the CBT units (Fig. 1b) and thus the formation of supramolecular nanofibres with molecular-scale diameter.

An important factor for efficient energy transport along the nanofibres is the electronic coupling between neighbouring CBT units. To determine this parameter, we resort to a reference compound, compound **2**, which has exactly the same supramolecular motif as compound **1**; however, the NIBT chromophores are replaced by octyl chains (Supplementary Fig. 5). The spectra of self-assembled compound **2** demonstrate the presence of H-aggregates. From those data we obtain a nearest-neighbour electronic coupling of $W \gtrsim 44$ meV ($350\ \mathrm{cm}^{-1}$; Supplementary Fig. 5), using the theoretical framework of ref. 27. We expect a similar value of W for compound **1**, as it has the same supramolecular motif. The magnitude of this electronic coupling between the CBT units is close to the strongest intermolecular coupling observed for other self-assembled nanostructures, such as J-aggregates based on small molecules[17,28] and photosynthetic light-harvesting antenna systems[29,30].

To study energy transport along single nanofibres, we diluted self-assembled compound **1** to a concentration of 0.07 μM (0.1 p.p.m., in *o*-DCB) and spin-coated this dispersion onto microscopy cover slips. Well isolated micrometre-long nanofibres were revealed by AFM (Fig. 2d, e and Supplementary Fig. 7). Owing to this large spatial separation, single nanofibres could be resolved and studied with a confocal microscope. We operated the microscope in imaging mode first, using widefield illumination and a charge-coupled-device (CCD) camera to detect the photoluminescence of the nanofibres. Figure 3a depicts a representative photoluminescence image, showing several individual nanofibres, in agreement with the AFM data (Supplementary Fig. 7). Having identified an isolated fibre (Fig. 3a, orange box), we switched the microscope to confocal illumination while the photoluminescence was still imaged on the CCD camera. We then positioned the nanofibre such that one of its ends coincided with the laser focus (which had a radius of ~300 nm; Fig. 3b, green spot). Intriguingly, photoluminescence from the entire structure over a distance of ~4 μm was observed. We rule out a waveguide effect, because the nanofibres are too narrow at 5 nm to produce such an effect; we also rule out direct photoexcitation more than about 500 nm away from the centre of the laser focus with control experiments on single molecules (as described in Supplementary Information section 7). Consequently, this photoluminescence signal must result from efficient transport of excitation energy over 4 μm. Given the typical π-stacking distance of 0.35 nm (refs 14, 15), transport over this distance must involve more than 10,000 molecules.

We investigated 97 individual fibres in total, to prove the robustness of this long-range energy transport. In most cases the transport distance is limited only by the nanofibre length (Fig. 3c and Supplementary Fig. 9). Such lengths, determined from photoluminescence images upon widefield illumination, range from 1.9 μm to 6.4 μm (average 3.3 μm). This range of lengths overlaps with the distribution of transport distances, between 1.6 μm and 4.4 μm (average 2.9 μm), retrieved from photoluminescence images upon confocal excitation at one fibre end (Supplementary Fig. 8 shows an example with interrupted energy transport).

To elucidate the energy-transport mechanism, we recorded photoluminescence spectra from single nanofibres (Supplementary Fig. 10). These data confirm that the photoluminescent NIBT periphery does not form structurally defined assemblies and therefore does not

Figure 2 | Characterization of self-assembled nanofibres. a, TEM image of a sample prepared from a gel of compound **1** (700 μM; 1,000 p.p.m.) in *ortho*-dichlorobenzene (*o*-DCB). **b**, Cross-section of the nanofibre from the yellow boxed area in **a**. Grey value is a measure of the intensity of the transmitted electron beam. Distance is the width of the yellow boxed area. **c**, Energy-minimized structure of compound **1** in its extended form. Dark grey, carbon atoms; light grey, hydrogen atoms; red, oxygen atoms; blue, nitrogen atoms; yellow, sulphur atoms. **d**, AFM image (topographical scan) of a spatially isolated single nanofibre, prepared by spin-coating a dispersion of self-assembled compound **1** (0.07 μM, 0.1 p.p.m., in *o*-DCB). **e**, Height profile along the dashed green arrow in **d**.

Figure 3 | Long-range energy transport along single supramolecular nanofibres. a, Widefield photoluminescence image of a spin-coated sample of a dispersion of self-assembled compound **1** (0.07 μM, 0.1 p.p.m., in *o*-DCB). The nanofibre in the boxed region appears slightly brighter than the other structures because the widefield illumination is not perfectly uniform. **b**, Photoluminescence image of the nanofibre from the boxed region in **a** upon confocal excitation of its bottom left end (filled green circle), demonstrating highly efficient energy transport over ~4 μm. **c**, Open bars show the distribution of the lengths of 97 fibres, determined from photoluminescence images upon widefield illumination; violet bars show the distribution of transport distances along single nanofibres retrieved upon confocal illumination of the same set of 97 nanofibres. **d**, Intensity profile along the orange dashed arrow in **b**. Distance refers to the length along this arrow. The emission intensity is normalized to its peak value. **e**, Illustration of the mechanism of energy transport along the nanofibre in **b**. Local illumination at one end (green arrow) gives rise to coherent energy transport along an ordered domain of the nanofibre's core (violet arrow). At small defects, here symbolized by a kink, incoherent energy transfer occurs either to the NIBT periphery, with subsequent photoluminescence (orange arrow), or to the next ordered domain of the core (black dashed arrow), whereupon coherent transport takes place to the nanofibre's end.

support transport of excitons over macroscopic distances[4,5]. Hence, the efficient long-range transport must be occurring along the ordered nanofibre core, promoted by substantial electronic coupling between the H-aggregated CBT units. The electronic coupling gives rise to the formation of vibronic singlet excitons with a small transition dipole moment for the lowest-energy transition (Supplementary Information section 4). This strongly reduces the rate of the main loss mechanism for electronic excitations from the CBT units of compound **1**, that is, energy transfer to the NIBT periphery[10]; this energy transfer to NIBT can thus no longer compete with transport along the core. However, excitation energy can be trapped at small defects within the core. Energy transfer to the periphery then becomes more likely and photoluminescence from NIBT is observed. In this sense, the NIBT emission reports on both the transport distance along the nanofibres and the structural order of the core. This interpretation is supported by

the spatially nonuniform photoluminescence intensity along the nanofibre upon confocal illumination (Fig. 3b, d). We attribute the photoluminescence maxima to small defects within the fibre core, where excitation energy leaks to the photoluminescent periphery. The smaller photoluminescence signal stems from the parts of the nanofibre where energy transfer to the periphery is less efficient because the core is highly ordered (Fig. 3e).

The remarkable transport distances of up to 4.4 μm along single nanofibres at room temperature demonstrate a high mobility of electronic excitations. Given the electronic coupling between CBT units (44 meV) and the excited-state lifetime of self-assembled reference compound **2** (2.3 ns), we estimate transport distances of between ~100 nm for diffusive (Förster-type) exciton hopping and ~8 μm for entirely coherent motion (Supplementary Information section 6). Exclusively incoherent hopping cannot account for our observations; however, fully coherent transport is also unlikely at room temperature[17,28]. We therefore suggest a combined coherent–incoherent motion, with a dominant coherent contribution. The electronic coupling promotes delocalization of electronic excitations over ordered domains along the core[22]—that is, a coherent sharing by many CBT units (coherent transport)—while between these domains incoherent hopping occurs (Fig. 3e). Such largely coherent long-range transport makes this system a promising candidate with which to develop new concepts for quantum information technologies and for efficient solar-energy conversion. For instance, H-type nanofibres will be useful for transporting energy in an efficient and directed way from a light-harvesting antenna system to a transducer for conversion into charge carriers. In addition, the strongly reduced transition dipole moment of the lowest-energy transition in H-aggregates may be beneficial for achieving a stable charge-separated state, because there is no competition with (super-radiant) emission.

1. Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F. & Buchleitner, A. Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ. Sci.* **5**, 9374–9393 (2012).
2. Laquai, F., Park, Y.-S., Kim, J.-J. & Basché, T. Excitation energy transfer in organic materials: from fundamentals to optoelectronic devices. *Macromol. Rapid Commun.* **30**, 1203–1231 (2009).
3. Siebbeles, L. D. A. & Grozema, F. C. (eds) *Charge and Exciton Transport Through Molecular Wires* (Wiley–VCH, 2011).
4. Menke, S. M. & Holmes, R. J. Exciton diffusion in organic photovoltaic cells. *Energy Environ. Sci.* **7**, 499–512 (2014).
5. Lin, J. D. A. *et al.* Systematic study of exciton diffusion length in organic semiconductors by six experimental methods. *Mater. Horiz.* **1**, 280–285 (2014).
6. Bolinger, J. C., Traub, M. C., Adachi, T. & Barbara, P. F. Ultralong-range polaron-induced quenching of excitons in isolated conjugated polymers. *Science* **331**, 565–567 (2011).
7. Vogelsang, J., Adachi, T., Brazard, J., Vanden Bout, D. A. & Barbara, P. F. Self-assembly of highly ordered conjugated polymer aggregates with long-range energy transfer. *Nature Mater.* **10**, 942–946 (2011).
8. Avakian, P. & Merrifield, R. E. Experimental determination of the diffusion length of triplet exciton in anthracene crystals. *Phys. Rev. Lett.* **13**, 541–543 (1964).
9. Dubin, F. *et al.* Macroscopic coherence of a single exciton state in an organic quantum wire. *Nature Phys.* **2**, 32–35 (2006).
10. Haedler, A. T. *et al.* Synthesis and photophysical properties of multichromophoric carbonyl-bridged triarylamines. *Chem. Eur. J.* **20**, 11708–11718 (2014).
11. Aida, T., Meijer, E. W. & Stupp, S. I. Functional supramolecular polymers. *Science* **335**, 813–817 (2012).
12. Seki, S., Saeki, A., Sakurai, T. & Sakamaki, D. Charge carrier mobility in organic molecular materials probed by electromagnetic waves. *Phys. Chem. Chem. Phys.* **16**, 11093–11113 (2014).
13. Sengupta, S. & Würthner, F. Chlorophyll J-aggregates: from bioinspired dye stacks to nanotubes, liquid crystals, and biosupramolecular electronics. *Acc. Chem. Res.* **46**, 2498–2512 (2013).
14. Cantekin, S., de Greef, T. F. A. & Palmans, A. R. A. Benzene-1,3,5-tricarboxamide: a versatile ordering moiety for supramolecular chemistry. *Chem. Soc. Rev.* **41**, 6125–6137 (2012).
15. Dou, X., Pisula, W., Wu, J., Bodwell, G. J. & Müllen, K. Reinforced self-assembly of hexa-peri-hexabenzocoronenes by hydrogen bonds: from microscopic aggregates to macroscopic fluorescent organogels. *Chem. Eur. J.* **14**, 240–249 (2008).

16. Scheibe, G., Schöntag, A. & Katheder, F. Fluoreszenz und Energiefortleitung bei reversibel polymerisierten Farbstoffen. *Naturwissenschaften* **29,** 499–501 (1939).
17. Eisele, D. M., Knoester, J. Kirstein, S., Rabe, J. P. & Vanden Bout, D. A. Uniform exciton fluorescence from individual molecular nanotubes immobilized on solid substrates. *Nature Nanotechnol.* **4,** 658–663 (2009).
18. Clark, K. A., Krueger, E. L. & Vanden Bout, D. A. Direct measurement of energy migration in supramolecular carbocyanine dye nanotubes. *J. Phys. Chem. Lett.* **5,** 2274–2282 (2014).
19. Lin, H. *et al.* Collective fluorescence blinking in linear J-aggregates assisted by long-distance exciton migration. *Nano Lett.* **10,** 620–626 (2010).
20. Zhang, W. *et al.* Supramolecular linear heterojunction composed of graphite-like semiconducting nanotubular segments. *Science* **334,** 340–343 (2011).
21. Winiger, C. B., Li, S., Kumar, G. R., Langenegger, S. M. & Häner, R. Long-distance electronic energy transfer in light-harvesting supramolecular polymers. *Angew. Chem. Int. Ed.* **53,** 13609–13613 (2014).
22. Eisele, D. M. *et al.* Robust excitons inhabit soft supramolecular nanotubes. *Proc. Natl Acad. Sci. USA* **111,** E3367–E3375 (2014).
23. Kasha, M., Rawls, H. R. & Ashraf El-Bayoumi, M. The exciton model in molecular spectroscopy. *Pure Appl. Chem.* **11,** 371–392 (1965).
24. Chaudhuri, D. *et al.* Enhancing long-range exciton guiding in molecular nanowires by H-aggregation lifetime engineering. *Nano Lett.* **11,** 488–492 (2011).
25. Field, J. E. & Venkataraman, D. Heterotriangulenes—structure and properties. *Chem. Mater.* **14,** 962–964 (2002).
26. Kivala, M. *et al.* Columnar self-assembly in electron-deficient heterotriangulenes. *Chem. Eur. J.* **19,** 8117–8128 (2013).
27. Spano, F. C. Modeling disorder in polymer aggregates: the optical spectroscopy of regioregular poly(3-hexylthiophene) thin films. *J. Chem. Phys.* **122,** 234701 (2005).
28. Scheblykin, I. G., Sliusarenko, O. Y., Lepnev, L. S., Vitukhnovsky, A. G. & Van der Auweraer, M. Strong nonmonotonous temperature dependence of exciton migration rate in J aggregates at temperatures from 5 to 300 K. *J. Phys. Chem. B* **104,** 10949–10951 (2000).
29. Cogdell, R. J., Gall, A. & Köhler, J. The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to *in vivo* membranes. *Q. Rev. Biophys.* **39,** 227–324 (2006).
30. Oostergetel, G. T., van Amerongen, H. & Boekema, E. J. The chlorosome: a prototype for efficient light harvesting in photosynthesis. *Photosynth. Res.* **104,** 245–255 (2010).

Author Contributions A.T.H., K.K. and H.-W.S. designed and prepared compounds **1** and **2**, and investigated their self-assembly. M.K. and N.H. synthesized the functionalized CBT core as a building block for the synthesis of compounds **1** and **2**. R.H., A.I., B.W. and J.K. designed and performed optical experiments on single nanofibres. All authors contributed to discussion of the data and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H. (richard.hildner@uni-bayreuth.de) or H.-W.S. (hans-werner.schmidt@uni-bayreuth.de).

METHODS

Materials and sample preparation. The synthesis, purification and characterization of compound **1** and reference compound **2** are described in detail elsewhere[10]. Self-assembled nanofibres of compound **1** were prepared in *ortho*-dichlorobenzene (*o*-DCB). Molecularly dissolved solutions of compound **1** were obtained in 1,1,2,2-tetrachloroethane (see Supplementary Information). All solvents were used as received. The following procedures for sample preparation were used for the various characterization techniques:

TEM. Compound **1** was heated in *o*-DCB at a concentration of 700 μM (1,000 p.p.m., 0.1 wt%) until a clear orange solution was obtained. Stable gels formed upon cooling. Then a carbon-coated copper grid was dipped into the gel and *o*-DCB was removed with filter paper.

Fourier-transform infrared spectroscopy. Fourier-transform infrared spectroscopy measurements were conducted on compound **1** in *o*-DCB at a concentration of 14 mM (20,000 p.p.m., 2 wt%).

AFM. Compound **1** at a concentration of 7 μM (10 p.p.m.) in *o*-DCB was heated until a clear orange solution was obtained. After cooling to room temperature, compound **1** was allowed to self-assemble for 24 h before spin-coating the dispersion on microscopy cover slips (borosilicate glass; refractive index $n_e = 1.5255$ at 546.1 nm; measured thickness 150 μm; Menzel). To prepare samples with well isolated nanofibres, self-assembled compound **1** at a concentration of 70 μM (100 p.p.m.) in *o*-DCB was diluted to 0.07 μM (0.1 p.p.m.) and immediately spin-coated. Finally, all samples were dried under vacuum.

Optical imaging and spectroscopy of single nanofibres. Isolated nanofibres for optical imaging and spectroscopy were prepared using two procedures. First, self-assembled compound **1** at 7 μM (10 p.p.m.) in *o*-DCB was further diluted to 0.07 μM (0.1 p.p.m.) in a single step without further heat treatment. This dispersion was immediately spin-coated onto microscopy cover slips (Menzel; see above). Second, self-assembled compound **1** at 70 μM (100 p.p.m.) in *o*-DCB was diluted to 0.07 μM (0.1 p.p.m.) and immediately spin-coated. Finally, all samples were dried under vacuum.

TEM. TEM images were recorded on a LEO 922 Omega electron microscope operated at 200 kV in bright-field mode.

Fourier-transform infrared spectroscopy. Fourier-transform infrared spectroscopy was performed on a Digilab Division FTS-40 spectrometer using a home-built sample cell with ZnS windows and 0.5 mm path length. For the measurement, a hot solution of compound **1** in *o*-DCB (14 mM, 20,000 p.p.m.) was injected into the preheated sample cell (160 °C). The heating was turned off and the sample was allowed to slowly cool to room temperature over ~60 min. Measurements were conducted during the cooling process and at room temperature.

Ultraviolet/visible and photoluminescence spectroscopy in solution. The ultraviolet/visible spectra were recorded on a JASCO V-670 spectrophotometer. The photoluminescence spectra were measured at room temperature on a JASCO FP-8600 spectrofluorometer, and the photoluminescence quantum efficiency was determined with an integrating sphere (ILF-835). Hellma QS quartz-glass was used as cuvettes. Depending on the concentrations, the path length of the cuvette was adapted (10 mm or 1 mm) to avoid optical densities above 2.

AFM. AFM images were recorded on a Dimension 3100 NanoScope V (Veeco Metrology Group). Scanning was performed in tapping mode using silicon nitride (Si$_3$N$_4$) cantilevers (OTESPA-R3, Bruker) with a typical spring constant of 26 N m^{-1} and a typical resonance frequency of 300 kHz. The AFM image in Supplementary Fig. 2 was taken with a Dimension Icon (Bruker) equipped with a Nano-Scope V controller. Scanning was performed in tapping mode using Si$_3$N$_4$ cantilevers (OMCL-AC160TS, Olympus) with a typical spring constant of 42 N m^{-1} and a typical resonance frequency of 300 kHz. Image processing and analysis was conducted with NanoScope Analysis V1.40 software. The discrepancy between the heights and the diameters of the nanofibres, as determined by AFM (2–2.5 nm) and TEM (5 nm), respectively, is a known phenomenon[17,20].

Optical imaging and spectroscopy of single nanofibres. Optical imaging and spectroscopy was performed using a home-built microscope[31,32]. The excitation source was a pulsed diode laser (LDH-P-C-450B, Picoquant; 20 MHz repetition rate, 70 ps pulse duration) that operates at a wavelength of 450 nm, at which both the self-assembled CBT and the NIBT chromophores absorb light. The laser light was spatially filtered and directed to the microscope which was equipped with an infinity-corrected high-numerical-aperture oil-immersion objective (PlanApo, 60×, numerical aperture 1.45; Olympus). The sample was placed in the focal plane of the objective, and the sample position was controlled by a piezo-stage (Tritor 102 SG, from piezosystem jena). Photoluminescence was collected by the same objective and passed a set of dielectric filters (dichroic beam splitter z460RDC, long-pass filter LP467; AHF Analysentechnik) to suppress scattered or reflected laser light.

In imaging mode, the photoluminescence signal was imaged onto a CCD camera (Orca-ER, Hamamatsu) by an objective lens. In this mode we used two illumination methods. First, for widefield illumination we flipped an additional lens (widefield lens) into the excitation beam path to focus the laser light into the back focal plane of the microscope objective. This allowed for nearly uniform illumination of a large area with ~70 μm diameter in the sample plane; however, the excitation intensity is slightly higher in the centre of the image. Thus, overview photoluminescence images of our samples can be acquired to identify single nanofibres (Fig. 3a). Second, for confocal illumination the widefield lens was removed and the laser light was tightly focused to a spot with a radius of ~300 nm in the sample plane (Supplementary Information, section 7). Because the photoluminescence was still imaged onto the CCD camera, we could visualize the spatial distribution of the photoluminescence signal from single nanofibres under local excitation conditions (Fig. 3b). We did not investigate single nanofibres with lengths shorter than 1.5 μm, in order to reduce the influence of the spatial extent of our confocal spot on the measured transport distances as much as possible.

In spectroscopy mode, we measured simultaneously the photoluminescence spectrum and the photoluminescence lifetime of single nanofibres using only confocal illumination. In this mode the detected photoluminescence stems exclusively from the illuminated area on a nanofibre. This photoluminescence signal was directed to a 70/30 beam-splitter cube. To record emission spectra, we focused 70% of the signal onto the entrance slit of a spectrograph (250IS, Bruker; 150 grooves per millimetre, blaze wavelength 500 nm) equipped with a back-illuminated electron-multiplying CCD camera (iXon DV887-BI, Andor Technology). For lifetime measurements, the remaining 30% of the photoluminescence signal was focused onto a single-photon-counting avalanche photodiode (MPD, Picoquant). The electrical signal of this detector was fed into a time-correlated single-photon-counting module (TimeHarp 200, Picoquant).

For all measurements the excitation intensities were 24 W cm^{-2} for confocal and 0.2 W cm^{-2} for widefield illumination. All experiments were carried out at room temperature under ambient conditions.

Molecular modelling. The energy-minimized structure of compound **1** in Fig. 2c was calculated using a free copy of Avogadro Version 1.1.0 with an MMFF94s force field.

31. Issac, A. *et al.* Single molecule studies of calix[4]arene-linked perylene bisimide dimers: relationship between blinking, lifetime and/or spectral fluctuations. *Phys. Chem. Chem. Phys.* **14,** 10789–10798 (2012).
32. Issac, A., Hildner, R., Hippius, C., Würthner, F. & Köhler, J. Stepwise decrease of fluorescence versus sequential photobleaching in a single multichromophoric system. *ACS Nano* **8,** 1708–1717 (2014).

LETTER

Long–range energy transport in single supramolecular nanofibres at room temperature

Andreas T. Haedler[1], Klaus Kreger[1], Abey Issac[2]†, Bernd Wittmann[2], Milan Kivala[3], Natalie Hammer[3], Jürgen Köhler[2], Hans–Werner Schmidt[1] & Richard Hildner[2]

Efficient transport of excitation energy over long distances is a key process in light-harvesting systems, as well as in molecular electronics[1–3]. However, in synthetic disordered organic materials, the exciton diffusion length is typically only around 10 nanometres (refs 4, 5), or about 50 nanometres in exceptional cases[6,7], a distance that is largely determined by the probability laws of incoherent exciton hopping. Only for highly ordered organic systems has the transport of excitation energy over macroscopic distances been reported—for example, for triplet excitons in anthracene single crystals at room temperature[8], as well as along single polydiacetylene chains embedded in their monomer crystalline matrix at cryogenic temperatures (at 10 kelvin, or −263 degrees Celsius)[9]. For supramolecular nanostructures, uniaxial long-range transport has not been demonstrated at room temperature. Here we show that individual self-assembled nanofibres with molecular-scale diameter efficiently transport singlet excitons at ambient conditions over more than four micrometres, a distance that is limited only by the fibre length. Our data suggest that this remarkable long-range transport is predominantly coherent. Such coherent long-range transport is achieved by one-dimensional self-assembly of supramolecular building blocks, based on carbonyl-bridged triarylamines[10], into well defined H-type aggregates (in which individual monomers are aligned cofacially) with substantial electronic interactions. These findings may facilitate the development of organic nanophotonic devices and quantum information technology.

Supramolecular chemistry uses directed noncovalent interactions between molecules to construct well defined architectures that provide functionalities beyond those of the constituent molecular building blocks[11–16]. For example, supramolecular nanostructures have been identified as potential components for transporting excitation energy in light-harvesting applications[17–22] such as solar cells. These nanostructures feature a controlled and spatially well defined arrangement of their building blocks, with substantial intermolecular electronic coupling, which is a requirement for efficient energy transport. The observed energy-migration distances of around 100 nm have been attributed to incoherent exciton hopping with some contribution of coherent motion—that is, a delocalization of electronic excitations over several building blocks. In those nanostructures J-type aggregates are formed, by a brick-layer-type arrangement of the constituent molecules[17–20,22]. In J-aggregates, however, the oscillator strength is redistributed into the lowest-energy exciton level, which forms a super-radiant state[23] with a much shorter excited-state lifetime with respect to that of the noninteracting building blocks. Consequently, the competition between radiative decay and energy transport strongly constrains the distance over which electronic excitations can migrate. In contrast, in ideal H-aggregates, with cofacially stacked building blocks, the lowest-energy transition is dipole forbidden—that is, optically not accessible[23]. This results in a strongly increased radiative lifetime for the relaxed excited state, which

should be beneficial for efficient energy transport over macroscopic distances[8,24].

Here we demonstrate, at room temperature, long-range energy transport along individual H-aggregated nanofibres of molecular-scale diameter. We use a specifically designed supramolecular building block (compound 1, Fig. 1a) that has a carbonyl-bridged triarylamine (CBT) as its core; this core is functionalized at positions 2, 6 and 10 via amide linkers with 4-(5-hexyl-2,2′-bithiophene)-naphthalimides (NIBTs)[10]. This combination of a planar, aromatic heterotriangulene

Figure 1 | Self-assembly of compound 1. Green, carbonyl-bridged triarylamine (CBT) core; blue, amide moieties; yellow/gold, 4-(5-hexyl-2,2′-bithiophene)-naphthalimide (NIBT) periphery. **a**, Chemical structure of compound **1**. **b**, Self-assembly into nanofibres with an ordered H-aggregated core, driven by π-stacking of CBTs and stabilized by three chains of hydrogen bonds between the amide groups.

[1]Macromolecular Chemistry I, Bayreuth Institute of Macromolecular Research, and Bayreuth Center for Colloids and Interfaces, University of Bayreuth, 95440 Bayreuth, Germany. [2]Experimental Physics IV and Bayreuth Institute of Macromolecular Research, University of Bayreuth, 95440 Bayreuth, Germany. [3]Chair of Organic Chemistry I, Department of Chemistry and Pharmacy, University of Erlangen–Nürnberg, 91054 Erlangen, Germany. †Present address: Department of Physics, Sultan Qaboos University, 123 Muscat, Oman.

core[25] and three hydrogen-bonding amide groups is the structure-defining element that enforces columnar self-assembly (Fig. 1b). The peripheral NIBT units exhibit a bright orange photoluminescence either upon direct photoexcitation or after energy transfer from CBT.

The pronounced self-assembly behaviour of compound **1** is demonstrated by gelation at a concentration of 700 μM (1,000 parts per million, p.p.m.) in *ortho*-dichlorobenzene (*o*-DCB)[10]. We used transmission electron microscopy (TEM) to study a solvent-free sample, directly prepared from the gel, and saw a dense network of nanofibres (Fig. 2a). Some fibres align next to each other, resulting in structures with widths of multiples of 5 nm (Supplementary Fig. 1). The mean diameter of single fibres of 5 ± 1 nm is close to the calculated diameter of 6 nm for compound **1** (Fig. 2b, c), revealing the presence of one-dimensional nanofibres with molecular-scale diameter. This finding is confirmed by the uniform fibre heights of 2–2.5 nm observed with atomic force microscopy (AFM) on samples spin-coated from a dispersion of self-assembled compound **1** in *o*-DCB (7 μM, 10 p.p.m.; Supplementary Fig. 2).

The self-assembly into nanofibres is driven by π-stacking of the aromatic CBT units (that is, by their cofacial arrangement driven by van der Waals interactions between their π-electron systems), as shown by changes in the optical spectra (Supplementary Fig. 3). In the absorption spectra, a strongly blueshifted band arises at about 380 nm (around 26,300 cm^{-1}) upon self-assembly in *o*-DCB, indicating the formation of H-aggregates. However, it is not clear which chromophores are involved in this process, because the absorption spectra of CBT and NIBT overlap[10]. In contrast, the photoluminescence of self-assembled compound **1** stems exclusively from the NIBT chromophores (Supplementary Fig. 3). The photoluminescence spectra do not feature new and strongly shifted bands that can be associated

with aggregation. Hence, the NIBT units are not stacked in an ordered fashion. Only the CBT units within the supramolecular nanofibres form well defined π-stacked assemblies. These one-dimensional H-aggregates are stabilized by intermolecular hydrogen bonds between the amide groups of compound **1**, as shown by changes in the N–H stretch vibration in Fourier-transform infrared spectra (Supplementary Fig. 4). Carbonyl-bridged triarylamines without amide groups are stacked with a clear offset[25,26]. In compound **1**, however, the three chains of hydrogen bonds enforce the cofacial arrangement of the CBT units (Fig. 1b) and thus the formation of supramolecular nanofibres with molecular-scale diameter.

An important factor for efficient energy transport along the nanofibres is the electronic coupling between neighbouring CBT units. To determine this parameter, we resort to a reference compound, compound **2**, which has exactly the same supramolecular motif as compound **1**; however, the NIBT chromophores are replaced by octyl chains (Supplementary Fig. 5). The spectra of self-assembled compound **2** demonstrate the presence of H-aggregates. From those data we obtain a nearest-neighbour electronic coupling of $W \gtrsim 44$ meV (350 cm^{-1}; Supplementary Fig. 5), using the theoretical framework of ref. 27. We expect a similar value of W for compound **1**, as it has the same supramolecular motif. The magnitude of this electronic coupling between the CBT units is close to the strongest intermolecular coupling observed for other self-assembled nanostructures, such as J-aggregates based on small molecules[17,28] and photosynthetic light-harvesting antenna systems[29,30].

To study energy transport along single nanofibres, we diluted self-assembled compound **1** to a concentration of 0.07 μM (0.1 p.p.m., in *o*-DCB) and spin-coated this dispersion onto microscopy cover slips. Well isolated micrometre-long nanofibres were revealed by AFM (Fig. 2d, e and Supplementary Fig. 7). Owing to this large spatial separation, single nanofibres could be resolved and studied with a confocal microscope. We operated the microscope in imaging mode first, using widefield illumination and a charge-coupled-device (CCD) camera to detect the photoluminescence of the nanofibres. Figure 3a depicts a representative photoluminescence image, showing several individual nanofibres, in agreement with the AFM data (Supplementary Fig. 7). Having identified an isolated fibre (Fig. 3a, orange box), we switched the microscope to confocal illumination while the photoluminescence was still imaged on the CCD camera. We then positioned the nanofibre such that one of its ends coincided with the laser focus (which had a radius of ~300 nm; Fig. 3b, green spot). Intriguingly, photoluminescence from the entire structure over a distance of ~4 μm was observed. We rule out a waveguide effect, because the nanofibres are too narrow at 5 nm to produce such an effect; we also rule out direct photoexcitation more than about 500 nm away from the centre of the laser focus with control experiments on single molecules (as described in Supplementary Information section 7). Consequently, this photoluminescence signal must result from efficient transport of excitation energy over 4 μm. Given the typical π-stacking distance of 0.35 nm (refs 14, 15), transport over this distance must involve more than 10,000 molecules.

We investigated 97 individual fibres in total, to prove the robustness of this long-range energy transport. In most cases the transport distance is limited only by the nanofibre length (Fig. 3c and Supplementary Fig. 9). Such lengths, determined from photoluminescence images upon widefield illumination, range from 1.9 μm to 6.4 μm (average 3.3 μm). This range of lengths overlaps with the distribution of transport distances, between 1.6 μm and 4.4 μm (average 2.9 μm), retrieved from photoluminescence images upon confocal excitation at one fibre end (Supplementary Fig. 8 shows an example with interrupted energy transport).

To elucidate the energy-transport mechanism, we recorded photoluminescence spectra from single nanofibres (Supplementary Fig. 10). These data confirm that the photoluminescent NIBT periphery does not form structurally defined assemblies and therefore does not

Figure 2 | Characterization of self-assembled nanofibres. a, TEM image of a sample prepared from a gel of compound **1** (700 μM; 1,000 p.p.m.) in *ortho*-dichlorobenzene (*o*-DCB). **b**, Cross-section of the nanofibre from the yellow boxed area in **a**. Grey value is a measure of the intensity of the transmitted electron beam. Distance is the width of the yellow boxed area. **c**, Energy-minimized structure of compound **1** in its extended form. Dark grey, carbon atoms; light grey, hydrogen atoms; red, oxygen atoms; blue, nitrogen atoms; yellow, sulphur atoms. **d**, AFM image (topographical scan) of a spatially isolated single nanofibre, prepared by spin-coating a dispersion of self-assembled compound **1** (0.07 μM, 0.1 p.p.m., in *o*-DCB). **e**, Height profile along the dashed green arrow in **d**.

Figure 3 | Long-range energy transport along single supramolecular nanofibres. a, Widefield photoluminescence image of a spin-coated sample of a dispersion of self-assembled compound **1** (0.07 μM, 0.1 p.p.m., in *o*-DCB). The nanofibre in the boxed region appears slightly brighter than the other structures because the widefield illumination is not perfectly uniform.
b, Photoluminescence image of the nanofibre from the boxed region in **a** upon confocal excitation of its bottom left end (filled green circle), demonstrating highly efficient energy transport over ~4 μm. **c**, Open bars show the distribution of the lengths of 97 fibres, determined from photoluminescence images upon widefield illumination; violet bars show the distribution of transport distances along single nanofibres retrieved upon confocal illumination of the same set of 97 nanofibres. **d**, Intensity profile along the orange dashed arrow in **b**. Distance refers to the length along this arrow. The emission intensity is normalized to its peak value. **e**, Illustration of the mechanism of energy transport along the nanofibre in **b**. Local illumination at one end (green arrow) gives rise to coherent energy transport along an ordered domain of the nanofibre's core (violet arrow). At small defects, here symbolized by a kink, incoherent energy transfer occurs either to the NIBT periphery, with subsequent photoluminescence (orange arrow), or to the next ordered domain of the core (black dashed arrow), whereupon coherent transport takes place to the nanofibre's end.

support transport of excitons over macroscopic distances[4,5]. Hence, the efficient long-range transport must be occurring along the ordered nanofibre core, promoted by substantial electronic coupling between the H-aggregated CBT units. The electronic coupling gives rise to the formation of vibronic singlet excitons with a small transition dipole moment for the lowest-energy transition (Supplementary Information section 4). This strongly reduces the rate of the main loss mechanism for electronic excitations from the CBT units of compound **1**, that is, energy transfer to the NIBT periphery[10]; this energy transfer to NIBT can thus no longer compete with transport along the core. However, excitation energy can be trapped at small defects within the core. Energy transfer to the periphery then becomes more likely and photoluminescence from NIBT is observed. In this sense, the NIBT emission reports on both the transport distance along the nanofibres and the structural order of the core. This interpretation is supported by

the spatially nonuniform photoluminescence intensity along the nanofibre upon confocal illumination (Fig. 3b, d). We attribute the photoluminescence maxima to small defects within the fibre core, where excitation energy leaks to the photoluminescent periphery. The smaller photoluminescence signal stems from the parts of the nanofibre where energy transfer to the periphery is less efficient because the core is highly ordered (Fig. 3e).

The remarkable transport distances of up to 4.4 μm along single nanofibres at room temperature demonstrate a high mobility of electronic excitations. Given the electronic coupling between CBT units (44 meV) and the excited-state lifetime of self-assembled reference compound **2** (2.3 ns), we estimate transport distances of between ~100 nm for diffusive (Förster-type) exciton hopping and ~8 μm for entirely coherent motion (Supplementary Information section 6). Exclusively incoherent hopping cannot account for our observations; however, fully coherent transport is also unlikely at room temperature[17,28]. We therefore suggest a combined coherent–incoherent motion, with a dominant coherent contribution. The electronic coupling promotes delocalization of electronic excitations over ordered domains along the core[22]—that is, a coherent sharing by many CBT units (coherent transport)—while between these domains incoherent hopping occurs (Fig. 3e). Such largely coherent long-range transport makes this system a promising candidate with which to develop new concepts for quantum information technologies and for efficient solar-energy conversion. For instance, H-type nanofibres will be useful for transporting energy in an efficient and directed way from a light-harvesting antenna system to a transducer for conversion into charge carriers. In addition, the strongly reduced transition dipole moment of the lowest-energy transition in H-aggregates may be beneficial for achieving a stable charge-separated state, because there is no competition with (super-radiant) emission.

1. Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F. & Buchleitner, A. Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ. Sci.* **5,** 9374–9393 (2012).
2. Laquai, F., Park, Y.-S., Kim, J.-J. & Basché, T. Excitation energy transfer in organic materials: from fundamentals to optoelectronic devices. *Macromol. Rapid Commun.* **30,** 1203–1231 (2009).
3. Siebbeles, L. D. A. & Grozema, F. C. (eds) *Charge and Exciton Transport Through Molecular Wires* (Wiley–VCH, 2011).
4. Menke, S. M. & Holmes, R. J. Exciton diffusion in organic photovoltaic cells. *Energy Environ. Sci.* **7,** 499–512 (2014).
5. Lin, J. D. A. *et al.* Systematic study of exciton diffusion length in organic semiconductors by six experimental methods. *Mater. Horiz.* **1,** 280–285 (2014).
6. Bolinger, J. C., Traub, M. C., Adachi, T. & Barbara, P. F. Ultralong-range polaron-induced quenching of excitons in isolated conjugated polymers. *Science* **331,** 565–567 (2011).
7. Vogelsang, J., Adachi, T., Brazard, J., Vanden Bout, D. A. & Barbara, P. F. Self-assembly of highly ordered conjugated polymer aggregates with long-range energy transfer. *Nature Mater.* **10,** 942–946 (2011).
8. Avakian, P. & Merrifield, R. E. Experimental determination of the diffusion length of triplet exciton in anthracene crystals. *Phys. Rev. Lett.* **13,** 541–543 (1964).
9. Dubin, F. *et al.* Macroscopic coherence of a single exciton state in an organic quantum wire. *Nature Phys.* **2,** 32–35 (2006).
10. Haedler, A. T. *et al.* Synthesis and photophysical properties of multichromophoric carbonyl-bridged triarylamines. *Chem. Eur. J.* **20,** 11708–11718 (2014).
11. Aida, T., Meijer, E. W. & Stupp, S. I. Functional supramolecular polymers. *Science* **335,** 813–817 (2012).
12. Seki, S., Saeki, A., Sakurai, T. & Sakamaki, D. Charge carrier mobility in organic molecular materials probed by electromagnetic waves. *Phys. Chem. Chem. Phys.* **16,** 11093–11113 (2014).
13. Sengupta, S. & Würthner, F. Chlorophyll J-aggregates: from bioinspired dye stacks to nanotubes, liquid crystals, and biosupramolecular electronics. *Acc. Chem. Res.* **46,** 2498–2512 (2013).
14. Cantekin, S., de Greef, T. F. A. & Palmans, A. R. A. Benzene-1,3,5-tricarboxamide: a versatile ordering moiety for supramolecular chemistry. *Chem. Soc. Rev.* **41,** 6125–6137 (2012).
15. Dou, X., Pisula, W., Wu, J., Bodwell, G. J. & Müllen, K. Reinforced self-assembly of hexa-peri-hexabenzocoronenes by hydrogen bonds: from microscopic aggregates to macroscopic fluorescent organogels. *Chem. Eur. J.* **14,** 240–249 (2008).

16. Scheibe, G., Schöntag, A. & Katheder, F. Fluoreszenz und Energiefortleitung bei reversibel polymerisierten Farbstoffen. *Naturwissenschaften* **29,** 499–501 (1939).
17. Eisele, D. M., Knoester, J. Kirstein, S., Rabe, J. P. & Vanden Bout, D. A. Uniform exciton fluorescence from individual molecular nanotubes immobilized on solid substrates. *Nature Nanotechnol.* **4,** 658–663 (2009).
18. Clark, K. A., Krueger, E. L. & Vanden Bout, D. A. Direct measurement of energy migration in supramolecular carbocyanine dye nanotubes. *J. Phys. Chem. Lett.* **5,** 2274–2282 (2014).
19. Lin, H. *et al.* Collective fluorescence blinking in linear J-aggregates assisted by long-distance exciton migration. *Nano Lett.* **10,** 620–626 (2010).
20. Zhang, W. *et al.* Supramolecular linear heterojunction composed of graphite-like semiconducting nanotubular segments. *Science* **334,** 340–343 (2011).
21. Winiger, C. B., Li, S., Kumar, G. R., Langenegger, S. M. & Häner, R. Long-distance electronic energy transfer in light-harvesting supramolecular polymers. *Angew. Chem. Int. Ed.* **53,** 13609–13613 (2014).
22. Eisele, D. M. *et al.* Robust excitons inhabit soft supramolecular nanotubes. *Proc. Natl Acad. Sci. USA* **111,** E3367–E3375 (2014).
23. Kasha, M., Rawls, H. R. & Ashraf El-Bayoumi, M. The exciton model in molecular spectroscopy. *Pure Appl. Chem.* **11,** 371–392 (1965).
24. Chaudhuri, D. *et al.* Enhancing long-range exciton guiding in molecular nanowires by H-aggregation lifetime engineering. *Nano Lett.* **11,** 488–492 (2011).
25. Field, J. E. & Venkataraman, D. Heterotriangulenes—structure and properties. *Chem. Mater.* **14,** 962–964 (2002).
26. Kivala, M. *et al.* Columnar self-assembly in electron-deficient heterotriangulenes. *Chem. Eur. J.* **19,** 8117–8128 (2013).
27. Spano, F. C. Modeling disorder in polymer aggregates: the optical spectroscopy of regioregular poly(3-hexylthiophene) thin films. *J. Chem. Phys.* **122,** 234701 (2005).
28. Scheblykin, I. G., Sliusarenko, O. Y., Lepnev, L. S., Vitukhnovsky, A. G. & Van der Auweraer, M. Strong nonmonotonous temperature dependence of exciton migration rate in J aggregates at temperatures from 5 to 300 K. *J. Phys. Chem. B* **104,** 10949–10951 (2000).
29. Cogdell, R. J., Gall, A. & Köhler, J. The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to *in vivo* membranes. *Q. Rev. Biophys.* **39,** 227–324 (2006).
30. Oostergetel, G. T., van Amerongen, H. & Boekema, E. J. The chlorosome: a prototype for efficient light harvesting in photosynthesis. *Photosynth. Res.* **104,** 245–255 (2010).

METHODS

Materials and sample preparation. The synthesis, purification and characterization of compound **1** and reference compound **2** are described in detail elsewhere[10]. Self-assembled nanofibres of compound **1** were prepared in *ortho*-dichlorobenzene (*o*-DCB). Molecularly dissolved solutions of compound **1** were obtained in 1,1,2,2-tetrachloroethane (see Supplementary Information). All solvents were used as received. The following procedures for sample preparation were used for the various characterization techniques:

TEM. Compound **1** was heated in *o*-DCB at a concentration of 700 μM (1,000 p.p.m., 0.1 wt%) until a clear orange solution was obtained. Stable gels formed upon cooling. Then a carbon-coated copper grid was dipped into the gel and *o*-DCB was removed with filter paper.

Fourier-transform infrared spectroscopy. Fourier-transform infrared spectroscopy measurements were conducted on compound **1** in *o*-DCB at a concentration of 14 mM (20,000 p.p.m., 2 wt%).

AFM. Compound **1** at a concentration of 7 μM (10 p.p.m.) in *o*-DCB was heated until a clear orange solution was obtained. After cooling to room temperature, compound **1** was allowed to self-assemble for 24 h before spin-coating the dispersion on microscopy cover slips (borosilicate glass; refractive index $n_e = 1.5255$ at 546.1 nm; measured thickness 150 μm; Menzel). To prepare samples with well isolated nanofibres, self-assembled compound **1** at a concentration of 70 μM (100 p.p.m.) in *o*-DCB was diluted to 0.07 μM (0.1 p.p.m.) and immediately spin-coated. Finally, all samples were dried under vacuum.

Optical imaging and spectroscopy of single nanofibres. Isolated nanofibres for optical imaging and spectroscopy were prepared using two procedures. First, self-assembled compound **1** at 7 μM (10 p.p.m.) in *o*-DCB was further diluted to 0.07 μM (0.1 p.p.m.) in a single step without further heat treatment. This dispersion was immediately spin-coated onto microscopy cover slips (Menzel; see above). Second, self-assembled compound **1** at 70 μM (100 p.p.m.) in *o*-DCB was diluted to 0.07 μM (0.1 p.p.m.) and immediately spin-coated. Finally, all samples were dried under vacuum.

TEM. TEM images were recorded on a LEO 922 Omega electron microscope operated at 200 kV in bright-field mode.

Fourier-transform infrared spectroscopy. Fourier-transform infrared spectroscopy was performed on a Digilab Division FTS-40 spectrometer using a home-built sample cell with ZnS windows and 0.5 mm path length. For the measurement, a hot solution of compound **1** in *o*-DCB (14 mM, 20,000 p.p.m.) was injected into the preheated sample cell (160 °C). The heating was turned off and the sample was allowed to slowly cool to room temperature over ~60 min. Measurements were conducted during the cooling process and at room temperature.

Ultraviolet/visible and photoluminescence spectroscopy in solution. The ultraviolet/visible spectra were recorded on a JASCO V-670 spectrophotometer. The photoluminescence spectra were measured at room temperature on a JASCO FP-8600 spectrofluorometer, and the photoluminescence quantum efficiency was determined with an integrating sphere (ILF-835). Hellma QS quartz-glass was used as cuvettes. Depending on the concentrations, the path length of the cuvette was adapted (10 mm or 1 mm) to avoid optical densities above 2.

AFM. AFM images were recorded on a Dimension 3100 NanoScope V (Veeco Metrology Group). Scanning was performed in tapping mode using silicon nitride (Si_3N_4) cantilevers (OTESPA-R3, Bruker) with a typical spring constant of 26 N m^{-1} and a typical resonance frequency of 300 kHz. The AFM image in Supplementary Fig. 2 was taken with a Dimension Icon (Bruker) equipped with a Nano-Scope V controller. Scanning was performed in tapping mode using Si_3N_4 cantilevers (OMCL-AC160TS, Olympus) with a typical spring constant of 42 N m^{-1} and a typical resonance frequency of 300 kHz. Image processing and analysis was conducted with NanoScope Analysis V1.40 software. The discrepancy

between the heights and the diameters of the nanofibres, as determined by AFM (2–2.5 nm) and TEM (5 nm), respectively, is a known phenomenon[17,20].

Optical imaging and spectroscopy of single nanofibres. Optical imaging and spectroscopy was performed using a home-built microscope[31,32]. The excitation source was a pulsed diode laser (LDH-P-C-450B, Picoquant; 20 MHz repetition rate, 70 ps pulse duration) that operates at a wavelength of 450 nm, at which both the self-assembled CBT and the NIBT chromophores absorb light. The laser light was spatially filtered and directed to the microscope which was equipped with an infinity-corrected high-numerical-aperture oil-immersion objective (PlanApo, 60×, numerical aperture 1.45; Olympus). The sample was placed in the focal plane of the objective, and the sample position was controlled by a piezo-stage (Tritor 102 SG, from piezosystem jena). Photoluminescence was collected by the same objective and passed a set of dielectric filters (dichroic beam splitter z460RDC, long-pass filter LP467; AHF Analysentechnik) to suppress scattered or reflected laser light.

In imaging mode, the photoluminescence signal was imaged onto a CCD camera (Orca-ER, Hamamatsu) by an objective lens. In this mode we used two illumination methods. First, for widefield illumination we flipped an additional lens (widefield lens) into the excitation beam path to focus the laser light into the back focal plane of the microscope objective. This allowed for nearly uniform illumination of a large area with ~70 μm diameter in the sample plane; however, the excitation intensity is slightly higher in the centre of the image. Thus, overview photoluminescence images of our samples can be acquired to identify single nanofibres (Fig. 3a). Second, for confocal illumination the widefield lens was removed and the laser light was tightly focused to a spot with a radius of ~300 nm in the sample plane (Supplementary Information, section 7). Because the photoluminescence was still imaged onto the CCD camera, we could visualize the spatial distribution of the photoluminescence signal from single nanofibres under local excitation conditions (Fig. 3b). We did not investigate single nanofibres with lengths shorter than 1.5 μm, in order to reduce the influence of the spatial extent of our confocal spot on the measured transport distances as much as possible.

In spectroscopy mode, we measured simultaneously the photoluminescence spectrum and the photoluminescence lifetime of single nanofibres using only confocal illumination. In this mode the detected photoluminescence stems exclusively from the illuminated area on a nanofibre. This photoluminescence signal was directed to a 70/30 beam-splitter cube. To record emission spectra, we focused 70% of the signal onto the entrance slit of a spectrograph (250IS, Bruker; 150 grooves per millimetre, blaze wavelength 500 nm) equipped with a back-illuminated electron-multiplying CCD camera (iXon DV887-BI, Andor Technology). For lifetime measurements, the remaining 30% of the photoluminescence signal was focused onto a single-photon-counting avalanche photodiode (MPD, Picoquant). The electrical signal of this detector was fed into a time-correlated single-photon-counting module (TimeHarp 200, Picoquant).

For all measurements the excitation intensities were 24 W cm^{-2} for confocal and 0.2 W cm^{-2} for widefield illumination. All experiments were carried out at room temperature under ambient conditions.

Molecular modelling. The energy-minimized structure of compound **1** in Fig. 2c was calculated using a free copy of Avogadro Version 1.1.0 with an MMFF94s force field.

31. Issac, A. *et al.* Single molecule studies of calix[4]arene-linked perylene bisimide dimers: relationship between blinking, lifetime and/or spectral fluctuations. *Phys. Chem. Chem. Phys.* **14,** 10789–10798 (2012).
32. Issac, A., Hildner, R., Hippius, C., Würthner, F. & Köhler, J. Stepwise decrease of fluorescence versus sequential photobleaching in a single multichromophoric system. *ACS Nano* **8,** 1708–1717 (2014).

LETTER

Basin–scale transport of hydrothermal dissolved metals across the South Pacific Ocean

Joseph A. Resing[1], Peter N. Sedwick[2], Christopher R. German[3], William J. Jenkins[3], James W. Moffett[4], Bettina M. Sohst[2] & Alessandro Tagliabue[5]

Hydrothermal venting along mid-ocean ridges exerts an important control on the chemical composition of sea water by serving as a major source or sink for a number of trace elements in the ocean[1–3]. Of these, iron has received considerable attention because of its role as an essential and often limiting nutrient for primary production in regions of the ocean that are of critical importance for the global carbon cycle[4]. It has been thought that most of the dissolved iron discharged by hydrothermal vents is lost from solution close to ridge-axis sources[2,5] and is thus of limited importance for ocean biogeochemistry[6]. This long-standing view is challenged by recent studies which suggest that stabilization of hydrothermal dissolved iron may facilitate its long-range oceanic transport[7–10]. Such transport has been subsequently inferred from spatially limited oceanographic observations[11–13]. Here we report data from the US GEOTRACES Eastern Pacific Zonal Transect (EPZT) that demonstrate lateral transport of hydrothermal dissolved iron, manganese, and aluminium from the southern East Pacific Rise (SEPR) several thousand kilometres westward across the South Pacific Ocean. Dissolved iron exhibits nearly conservative (that is, no loss from solution during transport and mixing) behaviour in this hydrothermal plume, implying a greater longevity in the deep ocean than previously assumed[6,14]. Based on our observations, we estimate a global hydrothermal dissolved iron input of three to four gigamoles per year to the ocean interior, which is more than fourfold higher than previous estimates[7,11,14]. Complementary simulations with a global-scale ocean biogeochemical model suggest that the observed transport of hydrothermal dissolved iron requires some means of physicochemical stabilization and indicate that hydrothermally derived iron sustains a large fraction of Southern Ocean export production.

Hydrothermal fluids are enriched in iron (Fe) and manganese (Mn) by more than 10^6 relative to ambient deep ocean concentrations[1], and corresponding gross hydrothermal fluxes to the oceans are probably greater than those from global riverine inputs[3]. However, it has been well documented that most of the hydrothermal Fe is lost from the dissolved phase in the vicinity of ridge-axis vents, where hot ($\sim350\,^\circ C$), acidic, anoxic hydrothermal fluids ascend and mix with cold, alkaline, oxic sea water, resulting in the formation of Fe-sulphides and/or Fe-oxyhydroxides[2,5], which are subsequently lost from solution owing to settling and scavenging. As a result of these removal processes, it has been widely assumed that seafloor hydrothermal emissions are not a major source of dissolved Fe (Fe_d) to the ocean[6]. In contrast, dissolved Mn (Mn_d) is oxidized more slowly than Fe_d in sea water, and hydrothermal Mn_d anomalies have been observed as far as 2,000 km from ridge-axis sources[15].

A number of recent studies have demonstrated that Fe_d can be stabilized against precipitation, aggregation, and scavenging losses from sea water by several different physicochemical mechanisms[7–10]. Such findings imply that hydrothermal activity could strongly affect the oceanic Fe_d inventory; however, comprehensive observational data on the persistence and fate of hydrothermal Fe are needed to evaluate this hypothesis. Although several recent studies have inferred the transport of hydrothermal Fe_d over distances of hundreds[16,17] to thousands of kilometres[11–13], those conclusions remain equivocal at the ocean basin-scale, owing to limited sampling coverage[11–13] and assumptions regarding synoptic distributions of the hydrothermal tracer helium-3 (^{3}He)[11,13,16].

Here we present data for samples collected from 35 hydrographic stations between Manta, Ecuador, and Papeete, Tahiti, during the US GEOTRACES Eastern Pacific Zonal Transect (GEOTRACES cruise GP16; Fig. 1). This expedition focused on the Peru upwelling region and the superfast-spreading southern East Pacific Rise (SEPR), one of the most volcanically active areas on Earth and the source of a well documented plume of hydrothermal ^{3}He that extends west across the deep South Pacific Ocean[18]. The data from this cruise reveal pronounced gradients in Fe_d, Mn_d, dissolved aluminium (Al_d), and excess ^{3}He ($^3He_{xs}$) concentrations along the $\sim$8,000-km-long cruise transect (Fig. 2).

The most striking and novel feature that we observed is a vast, mid-depth plume of elevated Fe_d and Mn_d that extends over a distance of more than 4,000 km to the west of the SEPR. This plume is carried by the westward-flowing mid-depth circulation[18–20], and is clearly defined by anomalous concentrations of $^3He_{xs}$ (Fig. 2). The distance over which Fe_d and Mn_d are transported from the SEPR is substantially greater than that observed in plumes identified from basin-scale sections across the Atlantic, Indian, Arctic, and Southern oceans[12,16,17,21–23]. Also notable are the elevated Al_d concentrations that extend more than 3,000 km west of the SEPR; enrichments of this magnitude and extent have not
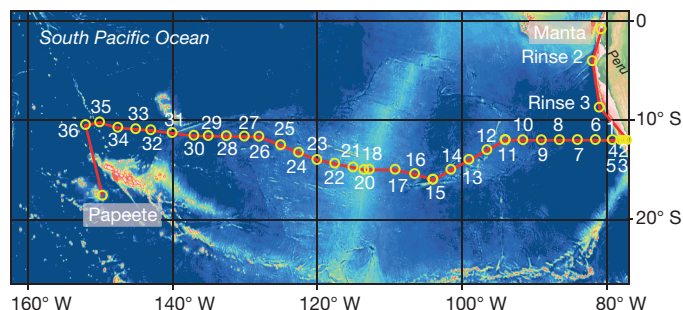
Figure 1 | Cruise track and station locations. The US GEOTRACES Eastern Pacific Zonal Transect (GEOTRACES cruise GP16) was undertaken on RV *Thomas G. Thompson* cruise 303 from 25 October to 20 December 2013. Station locations are shown as yellow circles with station numbers in white. Station 18 is located over the crest of the East Pacific Rise.

[1]Joint Institute for the Study of the Atmosphere and the Ocean, University of Washington and NOAA-PMEL, 7600 Sand Point Way NE, Seattle, Washington 98115, USA. [2]Department of Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, Virginia 23529, USA. [3]Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [4]Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway #AHF204, Los Angeles, California 90089, USA. [5]Department of Earth, Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, 4 Brownlow Street, Liverpool L69 3GP, UK.
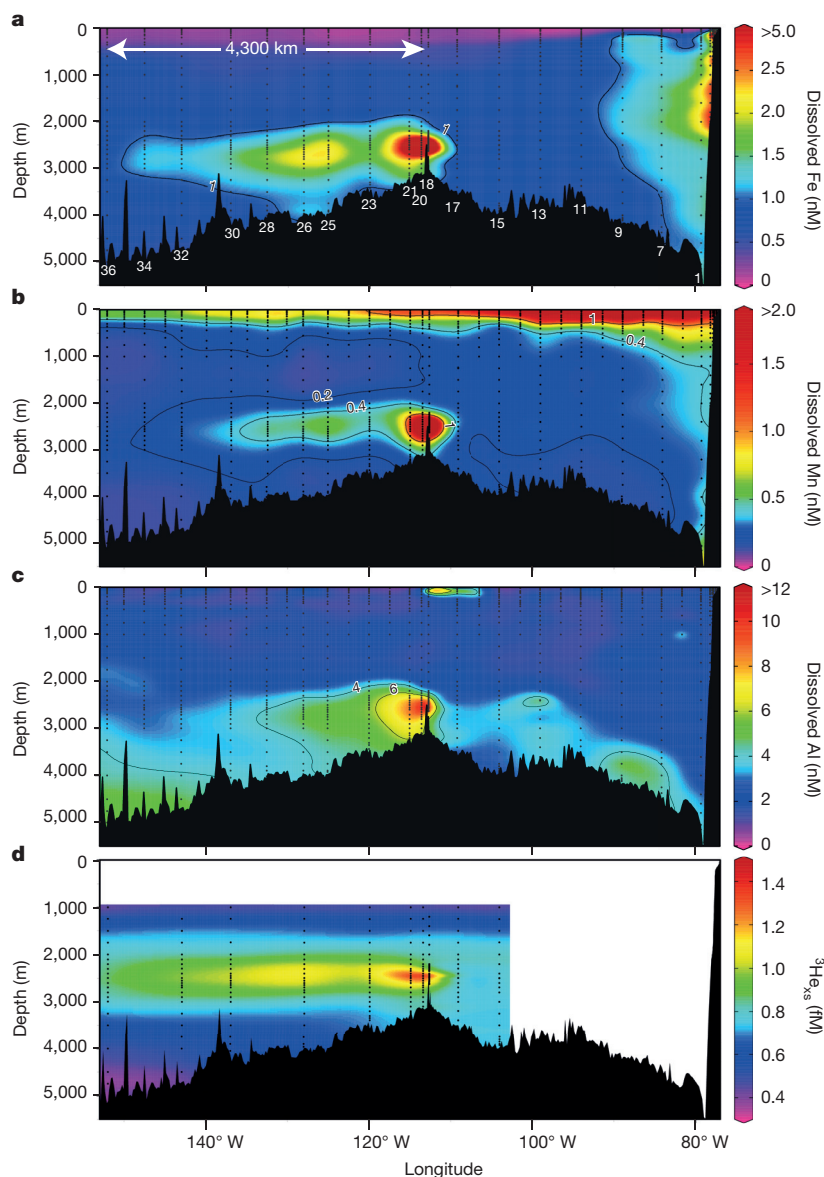
Figure 2 | Interpolated zonal concentration for GEOTRACES Eastern Pacific Zonal Transect. a, Dissolved iron. **b**, Dissolved manganese. **c**, Dissolved aluminium. **d**, Excess helium-3 ($^3He_{xs}$). Station numbers and distance west of East Pacific Rise are indicated on uppermost panel.

been previously reported for hydrothermal plumes in the Pacific, Arctic, Southern, or Indian oceans[12,21,22,24,25] (see also the GEOTRACES Intermediate Data Product, http://www.bodc.ac.uk/geotraces/data/idp2014/). Given the differences in geochemical behaviour between Fe and Mn, it is surprising that the lateral extent of the hydrothermal Fe_d anomaly exceeds that of Mn_d; inventories of hydrothermal Fe_d and Mn_d (Fe_d and Mn_d minus background) at station 32 are ~11% and ~4%, respectively, of those at station 20 (see Fig. 1). Our data set clearly documents the long-range transport of hydrothermal Fe_d from the SEPR, thus confirming the tentative conclusions drawn from limited previous observations[11,13].

Directly over the SEPR at station 18, Fe_d in the 'near field' hydrothermal plume is only ~20% of the total dissolvable Fe (an approximate measure of total hydrothermal Fe), and Fe(II) concentrations are near background levels. This suggests rapid oxidation and loss of hydrothermal Fe from the dissolved phase close to the ridge axis, consistent with previous observations from the SEPR[5]. In contrast, from the first off-axis station (station 20) continuing west across the basin as far as station 36, Fe_d concentrations are linearly correlated with $^3He_{xs}$ within the plume (Fig. 3a, b and Extended Data Fig. 1), indicating that hydrothermal Fe_d is behaving conservatively and therefore decreases in its concentration (as for the inert $^3He_{xs}$) reflect only mixing and dilution over a distance of ~4,300 km. Such behaviour is unexpected, given the

known propensity for the oxidation, aggregation, and scavenging of Fe_d from sea water[2,4,5]. Accordingly, our observations imply that Fe_d in the hydrothermal plume is somehow stabilized against loss from solution, perhaps as a result of complexation by dissolved organic ligands[7,8], or by incorporation into inorganic or organic colloids that reside within the dissolved (<0.2 μm) size fraction[9,10].

The relationship between Mn_d and $^3He_{xs}$ in the plume (Fig. 3c, d) indicates that hydrothermal Mn is removed from the dissolved phase until it reaches station 21, beyond which the residual hydrothermal Mn_d, like Fe_d, behaves conservatively with respect to $^3He_{xs}$. Dissolved Al over the ridge crest is enriched by as much as 12 nM over mid-depth concentrations to the east of the SEPR. This is comparable to enrichments in hydrothermal plumes over the Mid-Atlantic Ridge[23,26], where Al-rich plumes are spatially restricted to the deep axial valley and are thought to reflect the entrainment of Al-rich waters by rising hydrothermal fluids during plume formation[26]. In contrast, the SEPR typically lacks an axial valley, and the Al_d plume extends far from the ridge crest, suggesting a larger source of Al_d along the SEPR. Dissolved Al concentrations exceeding 100 nM have been reported in unusually acidic hydrothermal plumes that may be associated with seafloor eruptive activity[27], and the SEPR between 14° S and 19° S is a particularly active locus of seafloor volcanism, with hydrothermal and eruptive activity being more intense than along most other ridge sec-
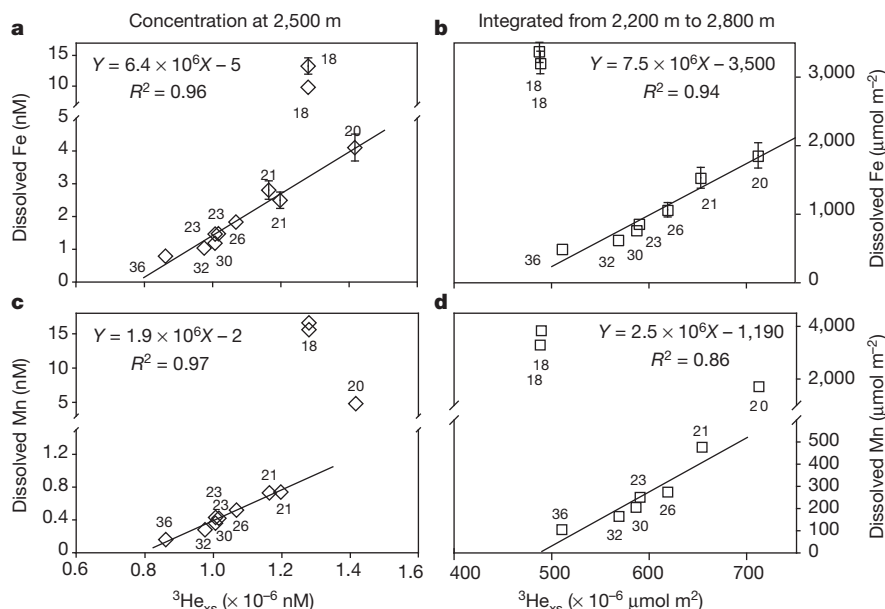
Figure 3 | Relationship between dissolved trace metals and ^{3}He west of SEPR. a, Dissolved iron versus ^{3}He$_{xs}$ at 2,500 m depth ($n = 11$). **b**, Dissolved Fe versus ^{3}He$_{xs}$, both integrated over a depth of 2,200–2,800 m except at station 18, where the maximum depth was 2,640 m ($n = 9$). **c**, Dissolved manganese versus ^{3}He$_{xs}$ at a depth of 2,500 m ($n = 11$). **d**, Dissolved manganese versus ^{3}He$_{xs}$ integrated as in **b** ($n = 9$). Error bars are twice the relative standard deviation of a given analysis, as reported in the Methods. Error bars are absent where the symbol size exceeds the error estimate. Lines represent the slope of a

simple linear regression analysis of the data. Discrete and integrated ^{3}He$_{xs}$ concentrations are lower at station 18 relative to stations west of the ridge; this difference is reduced for integrations between 2,200 m and 2,640 m depth (Extended Data Fig. 1a). The relatively low ^{3}He$_{xs}$ concentrations at station 18 ($\sim$15° S) suggest that the off-axis plume (stations 20–36) is primarily derived from vent fields located further south ($\sim$17° S–18.5° S) on the SEPR[5,28] with hydrothermal and eruptive effluent being homogenized and transported north and west[20] by along-axis and off-axis transport processes[20,30].

tions worldwide[5,28]. This suggests that eruptive activity is one possible source of the SEPR Al$_d$ plume.

We assess the importance of physicochemical stabilization to the long-range transport of hydrothermal Fe$_d$ using numerical simulations of Fe$_d$ and ^{3}He$_{xs}$ within a global-scale ocean biogeochemical

model that includes explicit cycling of dissolved Fe-binding ligands (see Methods). The model represents the input of ^{3}He as a function of ridge spreading rate and simulates hydrothermal Fe efflux via a fixed Fe:^{3}He ratio estimated from a global compilation of hydrothermal fluids[14]. Although there is reasonable qualitative agreement
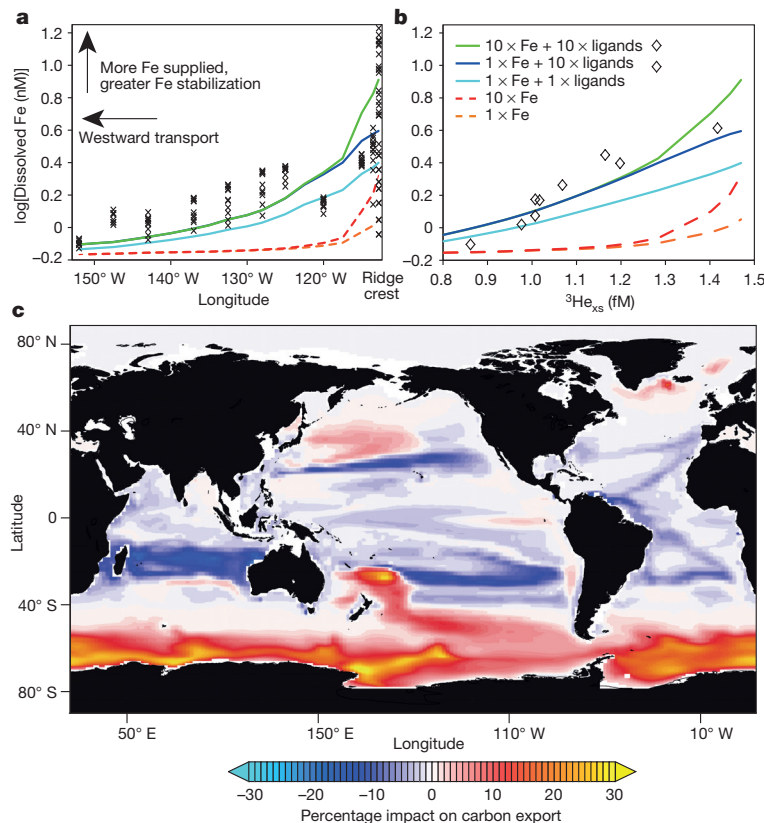
Figure 4 | Results of biogeochemical model simulations. Model simulation results centred at 2,530 m depth (average of two depth bins spanning 2,060–3,010 m; coloured lines) are shown in **a** and **b**. **a**, Dissolved Fe from model results compared to measured Fe$_d$ concentrations (crosses) between 2,200 m and 2,800 m to the west of the ridge axis. **b**, Dissolved Fe versus ^{3}He$_{xs}$ from model simulations compared to measured values (diamonds) at 2,500 m depth. For **a** and **b** the individual model simulations were run for 75 years. Orange dashed line, model solution using base hydrothermal Fe flux ($1 \times$ Fe). Red dashed line, $10 \times$ base hydrothermal Fe flux ($10 \times$ Fe). Cyan line, base hydrothermal Fe flux with equimolar flux of ligands ($1 \times$ Fe + $1 \times$ ligands). Dark blue line, base hydrothermal Fe flux with $10 \times$ greater ligand flux ($1 \times$ Fe + $10 \times$ ligands). Green line, $10 \times$ base Fe flux with equimolar flux of ligands ($10 \times$ Fe + $10 \times$ ligands). **c**, Percentage of annual export production due to hydrothermal Fe based on a 500-year model simulation employing base hydrothermal Fe flux with equimolar ligand flux ($1 \times$ Fe + $1 \times$ ligands) relative to a model solution with no hydrothermal Fe or ligand flux. Lower export production in the subtropical oceans is caused by decreased preformed macronutrients in the mode waters.

with our ^{3}He$_{xs}$ data (Extended Data Fig. 2), the sluggish deep-ocean circulation typical of relatively coarse-resolution global models restricts the overall westward transport of the hydrothermal anomaly (a feature that has proved challenging to measure and model[19,20]).

Even allowing for this insufficient abyssal propagation, Fe$_d$ concentrations and their relation to ^{3}He$_{xs}$ decrease rapidly to the west of the ridge crest owing to scavenging, when only hydrothermal Fe input is considered (Fig. 4a, b and Extended Data Fig. 3a); a tenfold increase in Fe input results in no improvement in the agreement with observations (Fig. 4a, b and Extended Data Fig. 3b). In contrast, when dissolved Fe-stabilizing 'ligands' from vent fluids or from processes occurring within the plume are added in an equimolar ratio with hydrothermal Fe$_d$, a much greater westward propagation of the Fe$_d$ plume is achieved and the model is better able to reproduce both the plume extent and relationship between Fe$_d$ and ^{3}He$_{xs}$ (Fig. 4a, b and Extended Data Fig. 3c–e). Adding tenfold more ligands, or tenfold more ligands and tenfold more hydrothermal Fe further increases the plume extent (note the logarithmic scale in Fig. 4a and b). Importantly, including a hydrothermal supply of ligands also improves the degree to which the model can reproduce the global distributions of Fe$_d$ in the abyssal ocean (Extended Data Table 1). A similar result would be expected for the formation of relatively unreactive Fe nanoparticles or colloids within the near-field plume[9,10], so the dissolved Fe-stabilizing 'ligands' could involve organic or inorganic moieties. These process-based model experiments indicate that the total input of hydrothermal Fe regulates the magnitude of the Fe$_d$ plume near the ridge crest, whereas the stabilization of Fe$_d$ against loss from solution governs its persistence and transport in the deep ocean.

Although recent global-scale models of the ocean Fe cycle suggest substantial hydrothermal contributions to the deep-ocean Fe inventory[14], our data indicate that this previous work substantially underestimated the far-field influence of hydrothermal Fe$_d$ emissions. The linear relationship between Fe$_d$ and ^{3}He$_{xs}$ concentrations in the plume west of the SEPR has a slope of $7.5 \pm 0.8 \times 10^6$ moles Fe$_d$ per mole ^{3}He$_{xs}$ (s.d. of slope based on simple linear regression; Fig. 3b), which falls roughly midway between values estimated for hydrothermal plumes in the western South Pacific, the Southern Ocean, and the South Atlantic[11,14,16]. If this relationship is representative of steady-state mid-ocean-ridge hydrothermal inputs to the ocean, then the estimated global hydrothermal ^{3}He efflux of 530 mol yr^{-1} (ref. 29) yields an 'effective' hydrothermal Fe$_d$ input of about 4 ± 1 Gmol yr^{-1} to the ocean interior, which is at least fourfold higher than previous estimates[7,11,14]. This Fe$_d$ is ultimately supplied to the iron-deficient surface waters of the Southern Ocean, where it supports ~15% to 30% of the modelled export production south of the Polar Front (Fig. 4c). The impact of hydrothermal Fe on export production is driven both by its gross flux and by processes that govern its stabilization (Extended Data Fig. 4). Thus, the ultimate impacts of hydrothermal activity on the biogeochemical cycle of Fe in the ocean may depend as much on the processes that control the longevity of hydrothermal Fe$_d$ plumes as on the magnitude of the hydrothermal Fe emissions, on which prior studies have largely focused.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 January; accepted 14 May 2015.

1. Von Damm, K. L. Seafloor hydrothermal activity: black smoker chemistry and chimneys. *Annu. Rev. Earth Planet. Sci.* **18,** 173–204 (1990).
2. German, C. R. & Seyfried, W. E. in *Treatise Geochemistry* Vol. 8 (eds Holland, H. D. & Turekian, K. K.) 191–233 (Elsevier, 2014).
3. Elderfield, H. & Schultz, A. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24,** 191–224 (1996).
4. Boyd, P. W. & Ellwood, M. J. The biogeochemical cycle of iron in the ocean. *Nature Geosci.* **3,** 675–682 (2010).
5. Feely, R. A. *et al.* Hydrothermal plume particles and dissolved phosphate over the superfast-spreading southern East Pacific Rise. *Geochim. Cosmochim. Acta* **60,** 2297–2323 (1996).
6. Bruland, K. & Lohan, M. in *Treatise Geochemistry* Vol. 6 (eds Holland, H. D. & Turekian, K. K.) 23–47 (Elsevier, 2003).
7. Bennett, S. A. *et al.* The distribution and stabilisation of dissolved Fe in deep-sea hydrothermal plumes. *Earth Planet. Sci. Lett.* **270,** 157–167 (2008).
8. Sander, S. G. & Koschinsky, A. Metal flux from hydrothermal vents increased by organic complexation. *Nature Geosci.* **4,** 145–150 (2011).
9. Yücel, M., Gartman, A., Chan, C. S. & Luther, G. W. Hydrothermal vents as a kinetically stable source of iron-sulphide-bearing nanoparticles to the ocean. *Nature Geosci.* **4,** 367–371 (2011).
10. Sands, C. M., Connelly, D. P., Statham, P. J. & German, C. R. Size fractionation of trace metals in the Edmond hydrothermal plume, Central Indian Ocean. *Earth Planet. Sci. Lett.* **319–320,** 15–22 (2012).
11. Fitzsimmons, J. N., Boyle, E. A. & Jenkins, W. J. Distal transport of dissolved hydrothermal iron in the deep South Pacific Ocean. *Proc. Natl Acad. Sci. USA* **111,** 16654–16661 (2014).
12. Nishioka, J., Obata, H. & Tsumune, D. Evidence of an extensive spread of hydrothermal dissolved iron in the Indian Ocean. *Earth Planet. Sci. Lett.* **361,** 26–33 (2013).
13. Wu, J., Wells, M. L. & Rember, R. Dissolved iron anomaly in the deep tropical–subtropical Pacific: evidence for long-range transport of hydrothermal iron. *Geochim. Cosmochim. Acta* **75,** 460–468 (2011).
14. Tagliabue, A. *et al.* Hydrothermal contribution to the oceanic dissolved iron inventory. *Nature Geosci.* **3,** 252–256 (2010).
15. Klinkhammer, G. & Hudson, A. Dispersal patterns for hydrothermal plumes in the South Pacific using manganese as a tracer. *Earth Planet. Sci. Lett.* **79,** 241–249 (1986).
16. Saito, M. A. *et al.* Slow-spreading submarine ridges in the South Atlantic as a significant oceanic iron source. *Nature Geosci.* **6,** 775–779 (2013).
17. Conway, T. M. & John, S. G. Quantification of dissolved iron sources to the North Atlantic Ocean. *Nature* **511,** 212–215 (2014).
18. Lupton, J. & Craig, H. A major helium-3 source at 15°S on the East Pacific Rise. *Science* **214,** 13–18 (1981).
19. Talley, L. & Johnson, G. Deep, zonal subequatorial currents. *Science* **263,** 1125–1128 (1994).
20. Hautala, S. & Riser, S. A nonconservative β-spiral determination of the deep circulation in the eastern South Pacific. *J. Phys. Oceanogr.* **23,** 1975–2000 (1993).
21. Middag, R., de Baar, H. J. W., Laan, P. & Bakker, K. Dissolved aluminium and the silicon cycle in the Arctic Ocean. *Mar. Chem.* **115,** 176–195 (2009).
22. Middag, R., van Slooten, C., de Baar, H. J. W. & Laan, P. Dissolved aluminium in the Southern Ocean. *Deep Sea Res. II* **58,** 2647–2660 (2011).
23. Measures, C., Hatta, M., Fitzsimmons, J. & Morton, P. Dissolved Al in the zonal N Atlantic section of the US GEOTRACES 2010/2011 cruises and the importance of Hydrothermal inputs. *Deep-Sea Res. II* **116,** 176–186 (2015).
24. Lupton, J. E. *et al.* Chemical and physical diversity of hydrothermal plumes along the East Pacific Rise, 8° 45′ N to 11° 50′ N. *Geophys. Res. Lett.* **20,** 2913–2916 (1993).
25. Klunder, M. B., Laan, P., Middag, R., de Baar, H. J. W. & van Ooijen, J. C. Dissolved iron in the Southern Ocean (Atlantic sector). *Deep Sea Res. II* **58,** 2678–2694 (2011).
26. Lunel, T., Rudnicki, M., Elderfield, H. & Hydes, D. Aluminium as a depth-sensitive tracer of entrainment in submarine hydrothermal plumes. *Nature* **344,** 137–139 (1990).
27. Gamo, T. *et al.* Hydrothermal plumes in the eastern Manus Basin, Bismarck Sea: CH$_4$, Mn, Al and pH anomalies. *Deep Sea Res. I* **40,** 2335–2349 (1993).
28. Urabe, T. *et al.* The effect of magmatic activity on hydrothermal venting along the superfast-spreading East Pacific Rise. *Science* **269,** 1092–1095 (1995).
29. Farley, K., Maier-Reimer, E., Schlosser, P. & Broeker, W. S. Constraints on mantle ^{3}He fluxes and deep-sea circulation from an oceanic general circulation model. *J. Geophys. Res.* **100,** 3829–3839 (1995).
30. Thurnherr, A. M., Ledwell, J. R., Lavelle, J. W. & Mullineaux, L. S. Hydrography and circulation near the crest of the East Pacific Rise between 9° and 10°N. *Deep. Sea Res. I* **58,** 365–376 (2011).

Author Contributions J.A.R. participated on the EPZT and determined Al$_d$ and Mn$_d$; P.N.S. interpreted the Fe$_d$ data; C.R.G. co-designed the study and participated in the EPZT; W.J.J. collected ^{3}He$_{xs}$ data; J.W.M. co-designed the study, participated in the EPZT, and collected Fe(II) data; B.M.S. participated in the EPZT and determined Fe$_d$; A.T. conducted the modelling experiments and interpreted their results. All authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.R. (resing@uw.edu) or A.T. (a.tagliabue@liverpool.ac.uk, for code).

METHODS

Sample collection and processing. Water column samples for trace metal analyses were obtained using 24 modified 12-litre GO-FLO bottles (General Oceanics) mounted on a trace-metal-clean conductivity–temperature–depth carousel (SeaBird) that was deployed on a Kevlar conducting cable[31]. Upon recovery, the GO-FLO samplers were brought into a shipboard Class-100 clean laboratory container for sub-sampling. For filtered samples, the samplers were pressurized to 10 psi using filtered, compressed air, and the seawater samples were filtered through pre-cleaned 0.2-µm Acropak Supor capsule filters (Pall) using rigorous trace-metal-clean protocols[31]. The Eastern Pacific Zonal Transect (EPZT) occupied 35 sampling stations along more than 8,000 km of cruise track. At 17 stations denoted as 'full' or 'super' stations, 37 samples were collected between the surface and the sea floor; at one additional station (station 34), 25 samples were collected over the upper 3,000 m of the water column. At 13 stations, denoted 'demi' stations, 13 samples were collected in the upper 1,000 m; and at 4 stations over the continental shelf, 7–24 samples were collected between the surface and sea floor. For Mn_d and Al_d, 876 samples were collected from all stations and depths. The 0.2-µm filtered subsamples were stored in 100-ml low-density polyethylene (LDPE) bottles (Bel-Art) with LDPE caps and were acidified to pH ≈ 1.7 with 12 N ultrapure hydrochloric acid (Fisher Optima). For Fe_d, 760 samples were collected at the full and super stations, with 0.2-µm filtered subsamples stored in 125-ml LDPE bottles with polypropylene caps (Nalgene) and acidified to pH ≈ 1.7 with a 6 N solution of ultrapure hydrochloric acid (Fisher Optima). Unfiltered seawater samples for the analysis of total dissolvable iron and manganese were collected at all full stations west of 109° W from the 24 deepest samples. These samples were collected in 125-ml LDPE bottles with polypropylene caps (Nalgene) and acidified to pH ≈ 1.7 with 12 N ultrapure hydrochloric acid (Fisher Optima), then stored for >4 months before analysis. Sample collection for Fe(II) analysis was identical to that for total dissolved metals, with the exception that the seawater samples were drawn into acid-washed 50-ml AirTite All-Plastic Norm-Ject syringes (Fisher Scientific) in the GEOTRACES sampling van to exclude oxygen contamination. These samples were stored on ice and in darkness to slow oxidation before analysis. Comparison with samples where Fe(II) was stabilized using 3-(N-morpholino)-propansulfonic acid buffer[32] indicated that Fe(II) was effectively preserved using the syringe protocol. Independent measurements of Fe_d indicated no detectable contamination from the syringes. Seawater samples for dissolved helium analysis (~45 g each) were drawn from the standard rosette (12-position, 30-litre Niskin-type bottles) using Tygon tubing connected to lengths of 5/8″ soft copper refrigeration tubing. Sample tubes were then hydraulically crimp-sealed[33].

Analytical methods. Dissolved Fe was determined at sea or at Old Dominion University by flow injection analysis with in-line pre-concentration on resin-immobilized 8-hydroxyquinoline and colorimetric detection[34,35], using a method modified from ref. 36. For the lowest-concentration samples from each analysis, and for SAFe seawater reference material S, the method of standard additions was used; all other samples were quantified using a standard curve obtained by addition of Fe standard solution to low-Fe sea water. For the cruise period, we determined the following Fe_d concentrations for the SAFe seawater reference materials: 0.126 ± 0.023 nM (n = 4) for SAFe seawater reference material S, and 1.26 ± 0.20 nM (n = 10) for SAFe seawater reference material D2. These values compare well with community consensus concentrations of 0.095 ± 0.008 nM and 0.955 ± 0.024 nM, respectively. In an effort to correct for day-to-day variations in analytical accuracy, all daily analyses included analysis of the GEOTRACES reference sea water, GSP, for which there is currently no consensus Fe_d concentration; all daily sample determinations were corrected using the difference between each day's measured GSP concentration and the overall cruise average Fe_d concentration for the GSP seawater (0.34 ± 0.07 nM, n = 27). The analytical limit of detection is estimated as the Fe_d concentration equivalent to a peak area that is three times the standard deviation on the 'zero-loading blank' ('manifold blank'), from which we estimate a detection limit of less than 0.04 nM[34,37]. Blank contributions from the ammonium acetate sample buffer solution (added on-line during analysis) and hydrochloric acid (added after collection) are negligible (that is, too low to quantify). Robust estimates of analytical precision are derived from multiple separate determinations of the SAFe seawater reference materials, which yield analytical uncertainties (expressed as one relative sample standard deviation on the mean) of ~15% at the concentration level of SAFe S (~0.1 nM) and ~10% at the concentration level of SAFe D2 (~1 nM). For high Fe (>5 nM) samples, Fe was determined by modifying the flow injection method to include a sample loop, rather than a pre-concentration column, and by using deionized water acidified to pH ≈ 1.7 as a carrier in place of the acid eluent. This modified flow injection method had an analytical precision of ±4% or ±1.5 nM (whichever is greater). Suitable seawater reference materials were not available for these analyses.

Dissolved Mn was determined at sea by flow injection analysis with in-line pre-concentration on resin-immobilized 8-hydroxyquinoline and colorimetric

detection[38]. Daily precision of analysis was ±0.01 nM (one standard deviation) or 3.8%, whichever is larger, based on the reproducibility of analytical and internal standards. A conservative estimate of the limit of detection is 0.03 nM based on three times the daily precision of analysis, which is consistent with previous work[38]. Two internal reference standards were run over the 57 days of the cruise, with Mn_d concentrations of 0.42 ± 0.036 nM (±8.4%; n = 102) and 0.31 ± 0.041 nM (±13%, n = 69), respectively. The SAFe reference samples were analysed simultaneously during sample analysis with the following results: for SAFe S, 0.85 ± 0.026 nM (n = 27; consensus value 0.79 ± 0.06 nM); for SAFe D2, 0.40 ± 0.028 nM (n = 22; consensus value 0.35 ± 0.05); and for SAFe D1, 0.36 ± 0.026 nM (n = 31; no consensus value). Analytical uncertainty is expressed as ± one standard deviation.

Dissolved Al was determined at sea by flow injection analysis with in-line pre-concentration and fluorimetric detection[39]. Method modifications included replacing resin-immobilized 8-hydroxyquinoline with Toyopearl AF-Chelate 650M, and using acidified de-ionized water as the carrier instead of acidified seawater. Daily precision for repeat analysis of internal and primary standards was ±0.1 nM or 4.2%, whichever is larger. Two internal reference standards were run during the cruise, with Al_d concentrations of 1.76 ± 0.25 nM (±14%; n = 101) and 1.98 ± 0.07 nM (±3.4%; n = 75), respectively. The SAFe reference samples were analysed simultaneously during sample analysis: for SAFe S, 2.38 ± 0.14 nM (n = 26; consensus value 1.67 ± 0.10 nM); for SAFe D2, 1.63 ± 0.13 nM (n = 26; consensus value 1.03 ± 0.09 nM); and for SAFe D1, 1.26 ± 0.11 nM (n = 32; consensus value 0.62 ± 0.03 nM). The least-squares best fit between our shipboard determinations and the SAFe consensus values is: $Al_{shipboard} = 1.02Al_{SAFe} + 0.59$ nM ($r^2 = 0.99$). Analytical uncertainty is expressed as ± one standard deviation. In the past, our laboratory has produced Al_d determinations that were statistically indistinguishable from the SAFe consensus concentrations, suggesting that our shipboard analytical method includes a consistent, unidentified blank equivalent to ~0.6 nM Al_d. Our estimated limit of detection of 0.3 nM based on daily precision estimates is low and might more conservatively be estimated to be >0.6 nM. The anomalously high Al_d concentrations (3.7–29.5 nM) determined in samples collected from 20–150 m depth at stations near 109° W and 113° W are not readily explained by ancillary chemical and physical data from the cruise, although there is no apparent reason to suspect that these few samples were contaminated during collection, processing, or analysis.

Fe(II) was determined at sea using an automated flow injection analysis system (FeLume II, Waterville Analytical) employing a Luminol chemiluminescence detection system[32,40]. The FeLume system was fitted with a standard quartz flow cell and a Hamamatsu HC135 photon counter configured with the following settings: pump speed of 15 rpm; photon counter integration time of 200 ms; load time of 20 s. The mean of the last 50 data points was used to determine the signal. Detection limits were determined for surface samples where ferrous iron was negligible based on a standard 3σ evaluation of the baseline signal[32,40]. This yielded an estimated detection limit of 14 pmol l^{-1}.

Helium was determined ashore after gases from the samples were quantitatively extracted under a vacuum into liquid-nitrogen chilled ~25-ml aluminosilicate glass flasks and sealed before analysis. Sample processing on the mass spectrometer system included purification over SAES getters to remove reactive gases and use of cryogenics to separate the noble gases[41–43]. Sample integrity was evaluated using noble gas abundances (not reported here), and determined by quadrupole mass spectrometer with an accuracy of 0.1%–0.5%, depending on the gas. The helium abundance and isotope ratio (^{3}He/^{4}He) were determined using a branch-tube magnetic sector mass spectrometer to an accuracy of 0.15% or better as determined by reproducibility of standards and duplicate samples. The isotope ratio was referenced to an atmospheric standard. Excess ^{3}He is computed as an approximate measure of the non-atmospheric ^{3}He over saturation:

$$^3He_{xs} = (\delta^3He - \delta^{*3}He)/100 \times C[He] \times 1.384 \times 10^{-6}$$

where $\delta^3He = 100 \times (R_x/R_a - 1) \times 100\%$, R_x and R_a are the ^{3}He/^{4}He ratios of the sample and air (1.384×10^{-6}), respectively. $\delta^{*3}He$ is the helium isotope ratio anomaly in solubility equilibrium with the atmosphere, which is a weak function of temperature[44] and is about −1.8% for the data used here. The precision of ^{3}He$_{xs}$ is 0.5% at 1 fM. The precision of ^{3}He$_{xs}$ is more than tenfold better than that of either Fe_d or Mn_d, allowing the use of a type I linear regression when comparing Fe_d and Mn_d to ^{3}He$_{xs}$.

The PISCES biogeochemical model. The model employed in this study is currently the only global-scale version that considers hydrothermal input of iron and a dynamic representation of iron-binding ligands[45]. The PISCES model[46,47] is a relatively complex ocean general circulation and biogeochemistry model that includes two phytoplankton groups, two zooplankton grazers, five limiting nutrients (nitrate, phosphate, silicic acid, ammonium, and Fe) and two size classes of organic carbon particles, calcium carbonate and biogenic silica, which sink and are

remineralized differentially. Dissolved Fe is supplied to the ocean from dust, sediments, rivers, and hydrothermal vents[47] and is ultimately lost to the sediments. The Fe_d is subjected to scavenging/coagulation losses, which produce two size classes of particulate iron. The scavenging rate is computed in the model by calculating the amount of 'free' uncomplexed Fe_d (assuming a dynamic ligand concentration and conditional stability) and the resulting net rate of scavenging depends on the concentrations of each particle species. We also account for the loss of organically complexed colloidal iron via coagulation processes and consider contributions from turbulent and Brownian components. The colloidal fraction of Fe_d is calculated as a function of temperature, salinity, and pH[48,49]. Ligand dynamics are represented assuming sources from phytoplankton or zooplankton exudation and organic matter degradation, and sinks associated with photochemical degradation, colloidal coagulation, and variable bacterial consumption, all on a reactivity continuum[45]. The ligand stability constants vary according to: $pKFe'L = 17.27 - 1565.7/TK$, where TK is absolute temperature leading to $pKFe'L$ of 11.5 at 0 °C and 11.9 at 20 °C[45]. Phytoplankton Fe_d uptake is computed using a quota model, with overall growth limitation accounting for the Fe demand associated with photosynthesis, respiration, and nitrate uptake[50]. PISCES is coupled to the three-dimensional ocean general circulation model NEMO, which has a spatial resolution of 2° of longitude, $2 \times \cos(\text{latitude})$ that is enhanced to 0.5° at the Equator, and 31 vertical levels, with the first ten levels in the upper 100 m. Hydrothermal transport was mostly observed over vertical levels 25 (centred on 2,290 m, depth range 2,050–2,530 m) and 26 (centred on 2,770 m, depth range 2530–3,010 m), which were used in Fig. 4a and b.

For this study we conducted a range of different simulations with PISCES aimed at addressing the processes responsible for the longevity of the observed hydrothermal Fe_d plume and their potential impact on the carbon cycle. The standard input flux of iron from the mid-ocean ridge was calculated based on iron-to-^{3}He ratios in hydrothermal fluids and spreading rate[14]; note that the flux is not based on the Fe_d flux calculated here. First, we conducted a set of experimental simulations over 75 years (outlined in Fig. 4) to examine the plume extent. To assess the large-scale impact of hydrothermal Fe and 'hydrothermal ligands' on ocean biogeochemistry and productivity, we extended the run with the standard addition of hydrothermal Fe_d with ligands in a 1:1 molar ratio ($1 \times Fe + 1 \times$ ligands) over a period of 500 years and compared that to a 500-year model run in which no hydrothermal Fe or ligands were added. After 500 years, the yearly change in biogeochemical tracers was negligible. In the model, ligands decay with time (microbial decay). As a result, the addition of hydrothermal ligands does not lead to their unrealistic accumulation in the ocean, with the main anomaly decaying rapidly from the ridge crest (Extended Data Fig. 5). Overall, the addition of hydrothermal ligands in the '$1 \times Fe + 1 \times$ ligands' and '$1 \times Fe + 10 \times$ ligands' experiments increases the total ligand inventory from 1.18×10^9 mol to 1.35×10^9 mol. To isolate the effect of hydrothermal ligand supply, we conducted a simulation that added only hydrothermal ligands without hydrothermal Fe_d and compared it to an experiment in which no ligands were added (Extended Data Fig. 5). This experiment is probably not representative of the real ocean, because ligands that might be produced at or near hydrothermal vents would be saturated with the Fe_d supplied by hydrothermal vents. As a consequence this experiment releases ligands into the ocean with an extremely high capacity to complex Fe_d from other sources. In this extreme hypothetical case where unsaturated hydrothermal ligands are able to bind Fe_d from other sources, only a small (~5%) increase in export production in the Southern Ocean is observed.

We have conducted a statistical analysis of the model against the most recent compilation of Fe_d averaged onto the World Ocean Atlas grid. We note that it is challenging to quantitatively evaluate global-scale iron models because we are obliged to compare to localized point measurements rather than having an objective climatology such as those available for macronutrients (for example, World Ocean Atlas). Both the model and the observations were gridded onto $1° \times 1°$ grid with 33 vertical levels. After log-transforming over a depth of 2,000–5,500 m depth, there are 1,025 unique data comparisons. In the abyssal ocean (2,000–5,500 m depth) we find support for the conclusions drawn from the visual model–data comparison in Fig. 4. The correlation (Extended Data Table 1) increases markedly when a source of hydrothermal ligands is applied, relative to the model runs that do not add hydrothermal ligands ($1 \times Fe$ and $10 \times Fe$).

Code availability. The NEMO-PISCES model we use in this work is freely available (http://www.nemo-ocean.eu/) under the CeCILL free software licence (http://www.cecill.info/index.en.html). We used version 3.4 of the NEMO model and a modified version of the PISCES biogeochemical model. These modifications concern the representation of dynamic ligand cycling and this is not yet present in the freely available NEMO release but will be provided upon contacting A.T.

Hydrothermal plume inventory estimates. Depth-integrated metal inventories for depth intervals of interest were estimated by summing the product of the average concentration of samples from two sequential depths and the difference

in depth between those samples. Where duplicate samples were collected to provide overlap between hydrocast sampling, the average concentration of the samples from overlapping depths was used.

Ocean Data View parameters and adjustments. Ocean Data View (ODV; http://odv.awi.de/) was used to produce Fig. 2. ^{3}He$_{xs}$ concentration data were contoured using ODV's Diva gridding algorithm with a signal-to-noise ratio of 10. Dissolved iron concentration data were contoured using the Diva gridding algorithm with a signal-to-noise ratio of 6.5 with negative values suppressed. Dissolved manganese concentration data were contoured using the Diva gridding algorithm with a signal-to-noise ratio of 4 with negative gridded values suppressed. Dissolved aluminium concentration data were contoured using the Diva gridding algorithm with a signal-to-noise ratio of 11 with negative gridded values suppressed.

The extremely high concentrations of these species over the ridge crest (station 18) resulted in interpolated concentrations at station 17 that vastly exceeded the actual measured concentrations. To circumvent this contouring artefact, an artificial background station was inserted at 111.5° W, which is halfway between the ridge-crest station (station 18) and the first station to the east (station 17). This 'background station' duplicated the measured depth and concentration data for each species from station 17. The black sample location indicators for this artificial station were removed from Fig. 2.

Cruise track selection. The latitude for the western portion of the EPZT cruise was selected to follow the 'downstream' core of the hydrothermal ^{3}He plume close to latitude 15° S, as determined from observations of previous research expeditions in the eastern South Pacific Ocean (the GEOSECS, HELIOS, and WOCE programmes). Urabe et al.[28] surveyed the SEPR axis between 13.8° S and 18.6° S, finding the most intense hydrothermal plumes between 17° S and 18.5° S. These plumes were rich in particulate iron[5] and total dissolvable manganese[51]. The total dissolvable manganese concentrations at both 2,500 m depth and integrated over 2,200–2,800 m depth in the most intense plumes over the SEPR axis in 1993 were greater than corresponding values observed at any of our EPZT cruise stations. As the ^{3}He data are unpublished, we examine ^{3}He: total dissolvable manganese Mn along the ridge crest[28], which has the highest values between 17° S and 18.5° S, suggesting that both discreet and integrated ^{3}He concentrations were much higher in those plumes compared to the plume we sampled at 15° S.

31. Cutter, G. A. & Bruland, K. W. Rapid and noncontaminating sampling system for trace elements in global ocean surveys. *Limnol. Oceanogr. Methods* **10,** 425–436 (2012).
32. Kondo, Y. & Moffett, J. W. Dissolved Fe(II) in the Arabian Sea oxygen minimum zone and western tropical Indian Ocean during the inter-monsoon period. *Deep Res. I* **73,** 73–83 (2013).
33. Young, C. & Lupton, J. E. An ultratight fluid sampling system using cold-welded copper tubing. *Eos* **64,** 735 (1983).
34. Sedwick, P. N., Bowie, A. R. & Trull, T. W. Dissolved iron in the Australian sector of the Southern Ocean (CLIVAR SR3 section): meridional and seasonal trends. *Deep Sea Res. I* **55,** 911–925 (2008).
35. Sedwick, P. N. et al. Iron in the Sargasso Sea (Bermuda Atlantic Time-series Study region) during summer: eolian imprint, spatiotemporal variability, and ecological implications. *Glob. Biogeochem. Cycles* **19,** GB4006 (2005).
36. Measures, C. I., Yuan, J. & Resing, J. A. Determination of iron in seawater by flow injection analysis using in-line preconcentration and spectrophotometric detection. *Mar. Chem.* **50,** 3–12 (1995).
37. Bowie, A. R., Sedwick, P. N. & Worsfold, P. J. Analytical intercomparison between flow injection-chemiluminescence and flow injection-spectrophotometry for the determination of picomolar concentrations of iron in seawater. *Limnol. Oceanogr. Methods* **2,** 42–54 (2004).
38. Resing, J. A. & Mottl, M. J. Determination of manganese in seawater using flow injection analysis with on-line preconcentration and spectrophotometric detection. *Anal. Chem.* **64,** 2682–2687 (1992).
39. Resing, J. & Measures, C. Fluorometric determination of Al in seawater by flow injection analysis with in-line preconcentration. *Anal. Chem.* **66,** 4105–4111 (1994).
40. Moffett, J. W., Goepfert, T. G. & Naqvi, S. W. A. Reduced iron associated with secondary nitrite maxima in the Arabian Sea. *Deep Sea Res. I* **54,** 1341–1349 (2007).
41. Stanley, R. H. R., Baschek, B., Lott, D. E. & Jenkins, W. J. A new automated method for measuring noble gases and their isotopic ratios in water samples. *Geochem. Geophys. Geosyst.* **10,** Q05008 (2009).
42. Lott, D. E. Improvements in noble gas separation methodology: a nude cryogenic trap. *Geochem. Geophys. Geosyst.* **2,** 2001GC000202 (2001).
43. Lott, D. & Jenkins, W. J. An automated cryogenic charcoal trap system for helium isotope mass spectrometry. *Rev. Sci. Instrum.* **55,** 1982–1988 (1984).
44. Benson, B. B. & Krause, D. J. Isotopic fractionation of helium during solution: a probe for the liquid state. *J. Solution Chem.* **9,** 895–909 (1980).
45. Völker, B. M. & Tagliabue, A. Modeling organic iron-binding ligands in a three-dimensional biogeochemical ocean model. *Mar. Chem.* **173,** 67–77 (2015).
46. Aumont, O. & Bopp, L. Globalizing results from ocean in situ iron fertilization studies. *Glob. Biogeochem. Cycles* **20,** GB2017 (2006).

47. Tagliabue, A., Aumont, O. & Bopp, L. The impact of different external sources of iron on the global carbon cycle. *Geophys. Res. Lett.* **41,** 920–926 (2014).
48. Liu, X. & Millero, F. J. The solubility of iron in seawater. *Mar. Chem.* **77,** 43–54 (2002).
49. Liu, X. & Millero, F. J. The solubility of iron hydroxide in sodium chloride solutions. *Geochim. Cosmochim. Acta* **63,** 3487–3497 (1999).

50. Flynn, K. & Hipkin, C. Interactions between iron, light, ammonium, and nitrate: insights from the construction of a dynamic model of algal physiology. *J. Phycol.* **35,** 1171–1190 (1999).
51. Ishibashi, J. *et al.* Hydrothermal methane and manganese variation in the plume over the superfast-spreading southern East Pacific Rise. *Geochim. Cosmochim. Acta* **61,** 485–500 (1997).

a

b
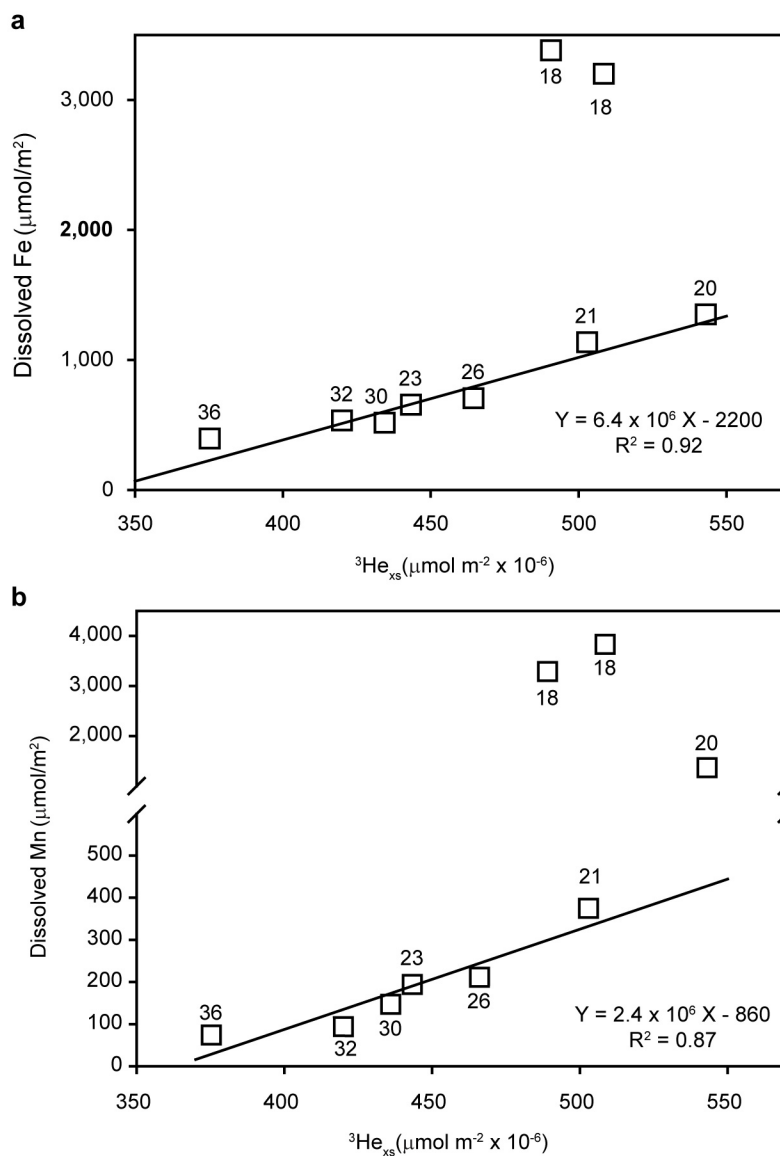
Extended Data Figure 1 | Relationship between dissolved trace metals and ^{3}HE. Depth-integrated concentrations of dissolved Fe (**a**), and dissolved Mn (**b**), versus depth-integrated concentration of ^{3}He$_{xs}$, over a depth range of 2,200–2,640 m. Sample station numbers are indicated for each data symbol.
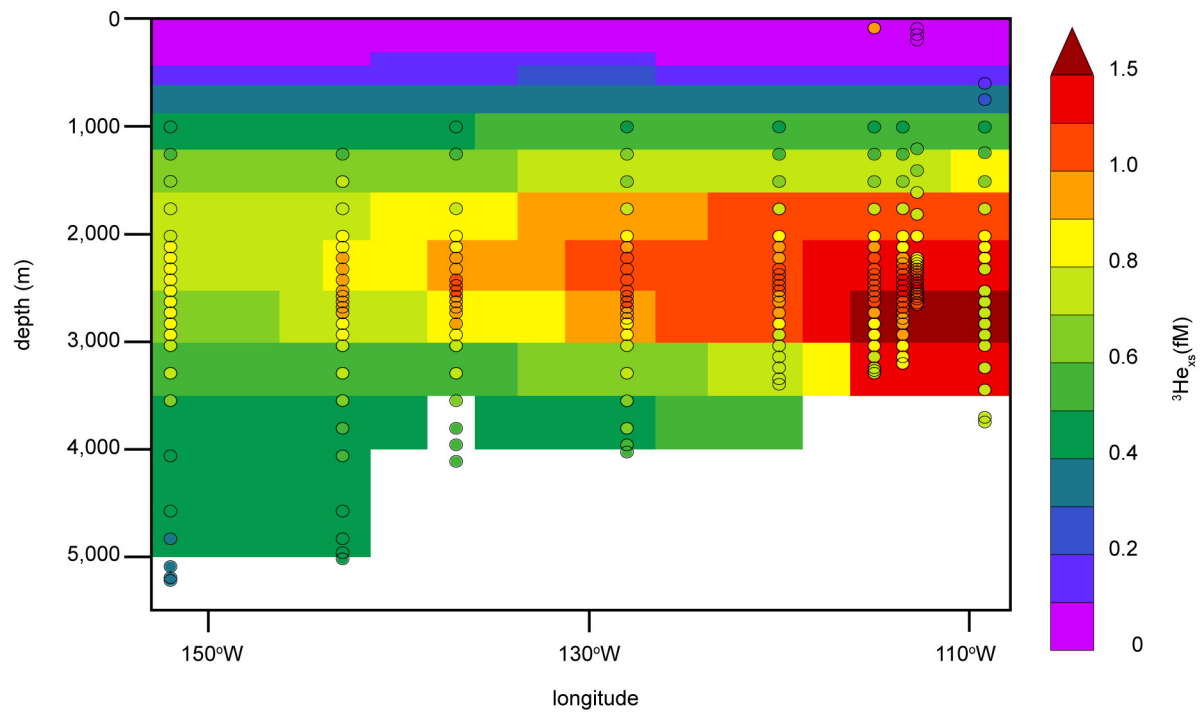
Extended Data Figure 2 | Comparison of modelled (rectangles) and measured (circular symbols) concentrations of $^3He_{xs}$, between EPZT cruise station 36 (far left) and station 17 (far right).
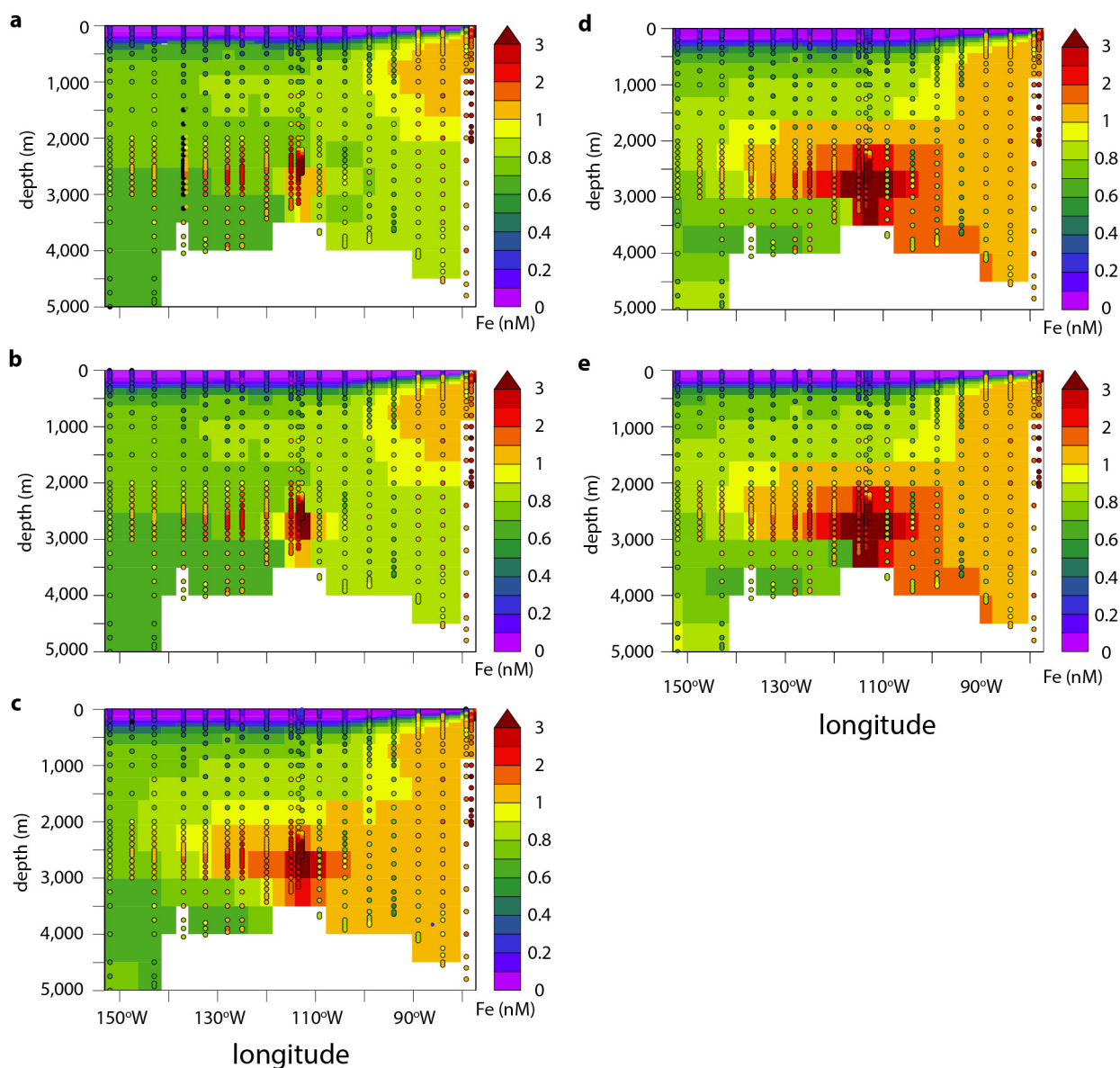
Extended Data Figure 3 | Sections of modelled Fe$_d$ transport and decay using a dynamic ligand global-circulation model[45] (see Methods). The model scenarios listed here are the same as those presented in Fig. 4. **a**, 1 × Fe; **b**, 10 × Fe; **c**, 1 × Fe + 1 × ligands; **d**, 1 × Fe + 10 × ligands; **e**, 10 × Fe + 10 × ligands.
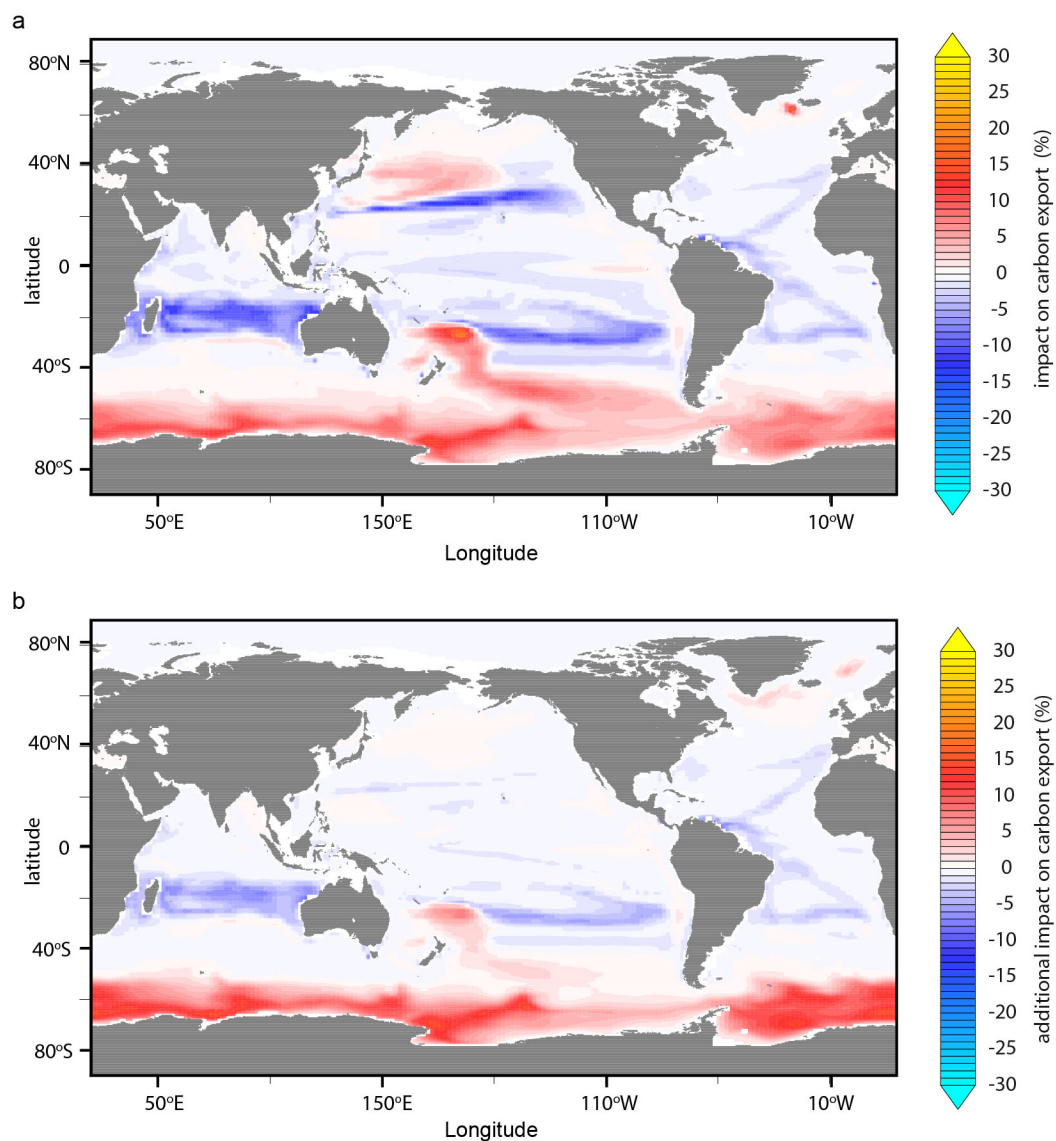
Extended Data Figure 4 | Impacts on carbon export from model simulations. **a**, Percentage contribution to carbon export production due to the input of hydrothermal Fe_d, not considering the addition of hydrothermal ligands. **b**, Additional percentage contribution from the addition of hydrothermal ligands to the simulation shown in **a**. **b** represents the difference between the total impact from the addition of both hydrothermal Fe_d and ligands (see Fig. 4c) compared to the input hydrothermal Fe_d without the addition of the ligands shown in **a**.

a No hydrothermal ligands

b 1X ligand flux (Experiment 3)
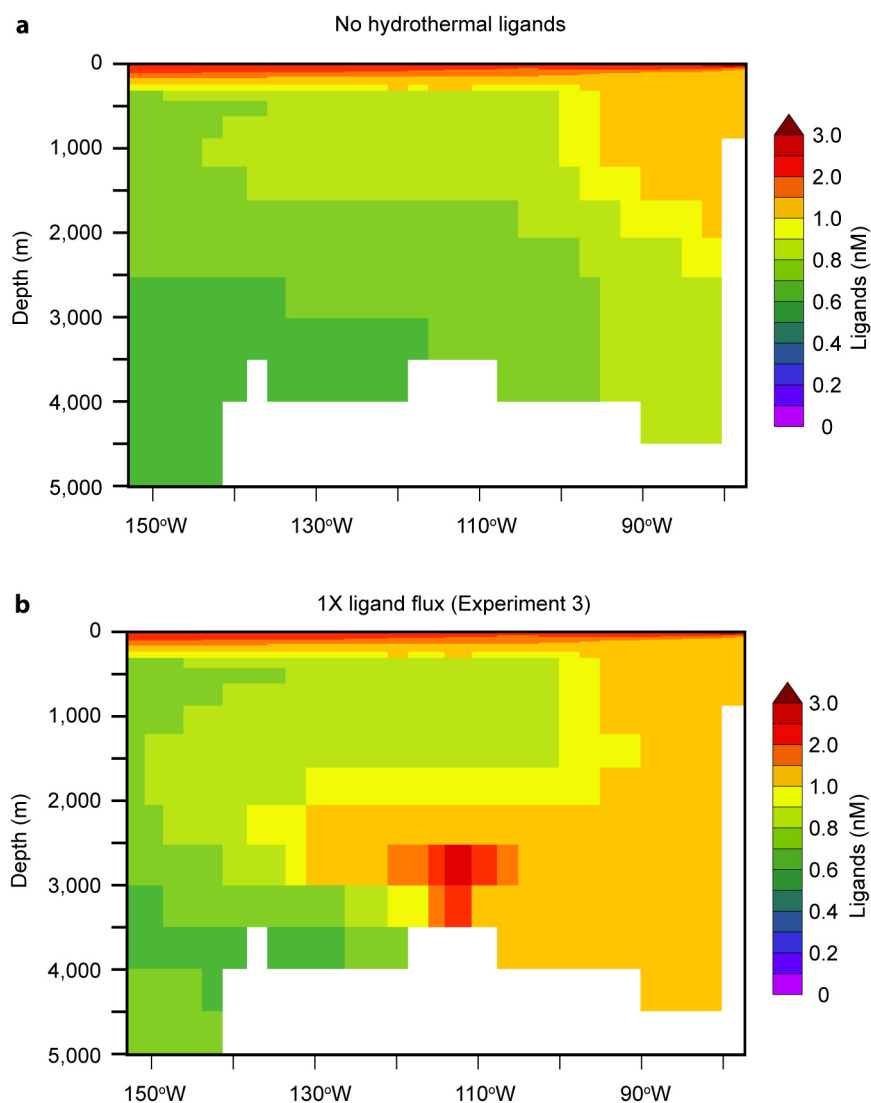
Extended Data Figure 5 │ Ligand flux model experiments. Two experiments were run to assess the impact of the flux of ligands associated with hydrothermal activity on the oceanic budget. **a**, Model simulation with no hydrothermal ligand flux. **b**, Model simulation with ligand flux equal to the flux of hydrothermal Fe.

Extended Data Table 1 | Model data comparison

		R	Mean Fe_d (nM)
Model scenarios	1 x Fe	0.12	0.59
	10 x Fe	0.30	0.61
	1 x Fe + 1 x ligands	0.42	0.71
	1 x Fe + 10 x ligands	0.47	0.87
	10 x Fe + 10 x ligands	0.59	1.02
Measurements	-	-	0.66

Comparisons between log-transformed model output and measurements of Fe_d averaged over the World Ocean Atlas grid ($1° \times 1°$ with 33 vertical levels) resulted in 1,025 individual model-data pairs at 2,000–5,500 m depth. R is the correlation between the modelled output and measurements.

LETTER

Eye–like ocelloids are built from different endosymbiotically acquired components

Gregory S. Gavelis[1], Shiho Hayakawa[1,2,3], Richard A. White III[4]†, Takashi Gojobori[3,5], Curtis A. Suttle[2,4,6,7], Patrick J. Keeling[2,7] & Brian S. Leander[1,2,7]

Multicellularity is often considered a prerequisite for morphological complexity, as seen in the camera-type eyes found in several groups of animals. A notable exception exists in single-celled eukaryotes called dinoflagellates, some of which have an eye-like 'ocelloid' consisting of subcellular analogues to a cornea, lens, iris, and retina[1]. These planktonic cells are uncultivated and rarely encountered in environmental samples, obscuring the function and evolutionary origin of the ocelloid. Here we show, using a combination of electron microscopy, tomography, isolated-organelle genomics, and single-cell genomics, that ocelloids are built from pre-existing organelles, including a cornea-like layer made of mitochondria and a retinal body made of anastomosing plastids. We find that the retinal body forms the central core of a network of peridinin-type plastids, which in dinoflagellates and their relatives originated through an ancient endosymbiosis with a red alga[2]. As such, the ocelloid is a chimaeric structure, incorporating organelles with different endosymbiotic histories. The anatomical complexity of single-celled organisms may be limited by the components available for differentiation, but the ocelloid shows that pre-existing organelles can be assembled into a structure so complex that it was initially mistaken for a multicellular eye[3]. Although mitochondria and plastids are acknowledged chiefly for their metabolic roles, they can also be building blocks for greater structural complexity.

Many organisms can orient to light. In some single-celled eukaryotes, such as *Chlamydomonas* and many dinoflagellates, an 'eyespot' directs photons onto photoreceptors on the flagellum, allowing the cell to respond to the intensity and direction of light[4,5]. A vastly more complex structure is found in warnowiid dinoflagellates: the eye-like ocelloid. Ocelloids consist of subcellular components resembling a lens, a cornea, iris-like rings, and a pigmented cup called the retinal body[6–9], which together so resemble the camera-type eyes of some animals that they have been speculated to be homologous[10] (Figs 1 and 2). The first description of a warnowiid was dismissed as a cell that had scavenged the eye from a jellyfish[3]. Ultrastructural studies of the ocelloid subsequently suggested that the retinal body might be derived from a plastid, in that it contains thylakoid-like membranes during cell division[4,8,9].

The ocelloid is among the most complex subcellular structures known, but its function and evolutionary relationship to other organelles remain unclear. This poor state of knowledge can be attributed to the fact that warnowiids are uncultivated and rarely encountered in environmental samples, with as few as two cells reported from the plankton per year for some species[11]. Modern single-cell genomics and microscopy approaches, however, provide opportunities to study uncultivated eukaryotes at the molecular and ultrastructural levels, including rare species[12–14]. In an attempt to learn more about the cell biology of ocelloids, we applied single-cell transcriptomics to two genera of warnowiids: *Erythropsidinium* (Supplementary Video 1) and *Warnowia* (Supplementary Video 2), as well as transmission electron

microscopy (TEM) on *Erythropsidinium* sp. and *Nematodinium* sp. Lastly, we investigated the three-dimensional ultrastructure and phylogenetic origin of the retinal body in *Nematodinium* sp. by using focused ion beam scanning electron microscopy (FIB-SEM) on isolated cells, and single-organelle genomics.

Thylakoid-like structures have been reported only once before in the retinal body[6], so we examined the ultrastructure of the ocelloid in *Nematodinium* sp. and *Erythropsidinium* sp. using single-cell TEM. During interphase, the retinal body contains highly ordered waveform membranes (Fig. 2), which are perpendicular to the plane expected for thylakoids in a chloroplast. However, we confirmed that near the end of interphase, the waveform membranes de-differentiated into a plastid-like arrangement made of double-stacked thylakoid-like structures (Extended Data Figs 1–3). Thus, the thylakoids and waveform membranes represent two modes of the same membrane system. Moreover, we found that the retinal body of *Nematodinium* sp. exhibits red fluorescence under 505 nm (green) light—suggesting the presence of chlorophyll or another autofluorescent pigment (Extended Data Fig. 4e). In *Nematodinium*, we also found mitochondria in the ocelloid, where they formed a cornea-like layer overlying the lens (Fig. 1c and Extended Data Fig. 5)[1].

To investigate further the possible plastid origin of the retinal body, we first examined transcriptomes from isolated cells of *Erythropsidinium* sp. and *Warnowia* sp., which appear to lack photosynthetic plastids. From polyadenylated complementary DNA (cDNA) libraries, we found that these heterotrophic genera expressed multiple photosynthesis-related genes (GenBank accession numbers KR632763–KR632773), including light-harvesting proteins. In addition, *Warnowia* sp. expressed three transcripts corresponding to the chloroplast-soluble peridinin–chlorophyll–binding protein, which is distinctive for dinoflagellate peridinin-type plastids[15].

The provenance of the retinal body is, however, concealed by the complex history of plastids in dinoflagellates[16]. While the ancestral peridinin-type plastid of dinoflagellates was initially acquired from a red alga, several dinoflagellates have since replaced this plastid with those from either a haptophyte, a cryptophyte, a diatom, or a green alga, and several non-photosynthetic lineages have been found to possess relict plastids[2,16,17]. To investigate the phylogenetic origin of the retinal body more directly, we characterized genes encoded on DNA associated with the organelle structures. Single cells of *Nematodinium* sp. were micro-dissected, and individual retinal bodies were isolated (Fig. 1). Retinal bodies were washed three times, as contaminant DNA can be a confounding factor in any genomic study. The individual retinal bodies from five cells were pooled, lysed, and their DNA was amplified with phi29 polymerase through multiple displacement amplification. To compare the DNA content of dissected organelles with the DNA content of whole *Nematodinium* cells (including nuclei), we

[1]Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [2]Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [3]Center for Information Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [4]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [5]Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. [6]Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [7]Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, Canada.
†Present address: Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99354, USA.
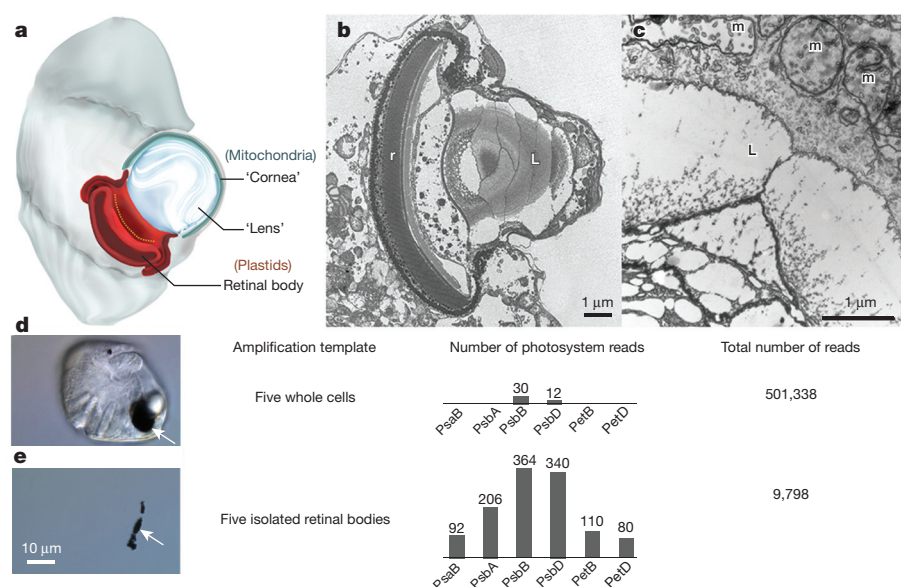
Figure 1 | Genomics and structure of organelles in the ocelloid. a, Illustration of *Nematodinium* showing the basic components of the ocelloid with their putative organellar origins. **b**, TEM of the ocelloid of *Erythropsidinium*, including the lens (L) and retinal body (r). **c**, TEM of the ocelloid of *Nematodinium*, depicting the edge of the lens (L) where it is overlain by a cornea-like layer of mitochondria (m). **d**, Genomic reads amplified from five whole cells of *Nematodinium*; arrow, retinal body. **e**, Genomic reads amplified from five retinal bodies (arrow) after they were micro-dissected from individual cells of *Nematodinium*.

also pooled five intact *Nematodinium* sp. cells and subjected them to the same procedures for DNA amplification and sequencing. From sequence databases derived from both samples, we identified genes that are encoded in the plastid of other dinoflagellates. Overall, six plastid genes were identified from isolated retinal bodies, *PsaB*, *PsbA*, *PsbB*, *PsbD*, *PetB*, and *PetD*, spanning photosystems I and II. These genes grouped strongly with the peridinin-containing plastids of dinoflagellates in individual and concatenated phylogenetic analysis (Fig. 3 and Extended Data Figs 6 and 7), and, collectively, plastid-encoded genes represented 13% of all reads. By contrast, the proportion of plastid/nuclear DNA in the whole-cell amplification was less than 0.0001%. The representation of plastid DNA in the retinal body was, therefore, over 1,600-fold higher than in whole cells (Fig. 1). While *in situ* hybridization is required to conclude firmly that plastid genomic DNA is localized within the retinal body, our findings strongly suggest that the retinal body is associated with a plastid genome.

Although the genomic data suggest that the retinal body is a derived plastid, there is another potential source of plastid DNA within the cell. Our isolates of *Nematodinium* contained small brown-pigmented bodies with double-stacked thylakoids typical of peridinin-type plastids. The presence of these plastids in addition to the retinal body raises

the possibility that *Nematodinium* has two different morphotypes of peridinin plastid within the same cell. However, the physical relationship between these plastid types was unclear from TEM alone, and the retinal body retains a distinct pigmentation as well as producing daughter retinal bodies through binary fission[6,8].

To investigate the physical connections between the different components of the ocelloid and surrounding structures, such as peridinin-type plastids, we performed FIB-SEM tomography on a single isolated cell of *Nematodinium* sp. The three-dimensional reconstructions of our FIB-SEM data demonstrated that the outer membrane of the retinal body is fused to a network of adjacent plastids, forming a membranous web throughout the cell (Fig. 4, Extended Data Fig. 8 and Supplementary Video 3). Therefore, the retinal body appears to be a differentiated region of a larger, netlike plastid. The fact that this plastid network was not evident in previous TEM-based studies of *Nematodinium*[18] suggests that hidden organelle networks could be widely overlooked in nature. Functional differentiation of discrete regions of plastids is known in other contexts, such as the pyrenoid—a centralized carbon-fixing region in many plastids—or the eyespots of some other eukaryotes, which consist of an intra-plastidial pigment cluster facing the flagellum[4,5].
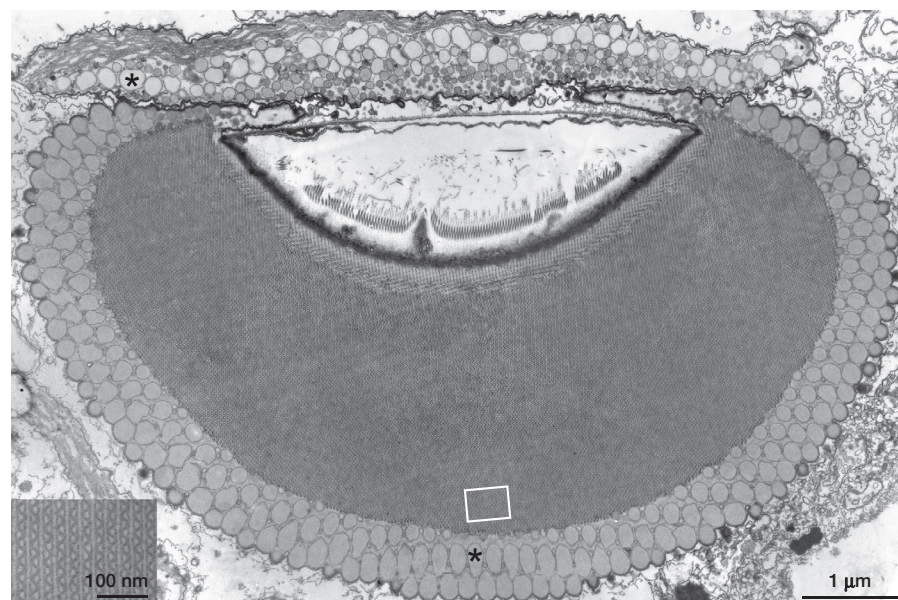
Figure 2 | Ultrastructure of the retinal body in *Nematodinium* sp. A composite of 12 electron micrographs showing a glancing section through the retinal body, which contains stacked waveform membranes (white square and inset) enveloped by pigmented lipid droplets (asterisk). Scale bar, 1 μm.
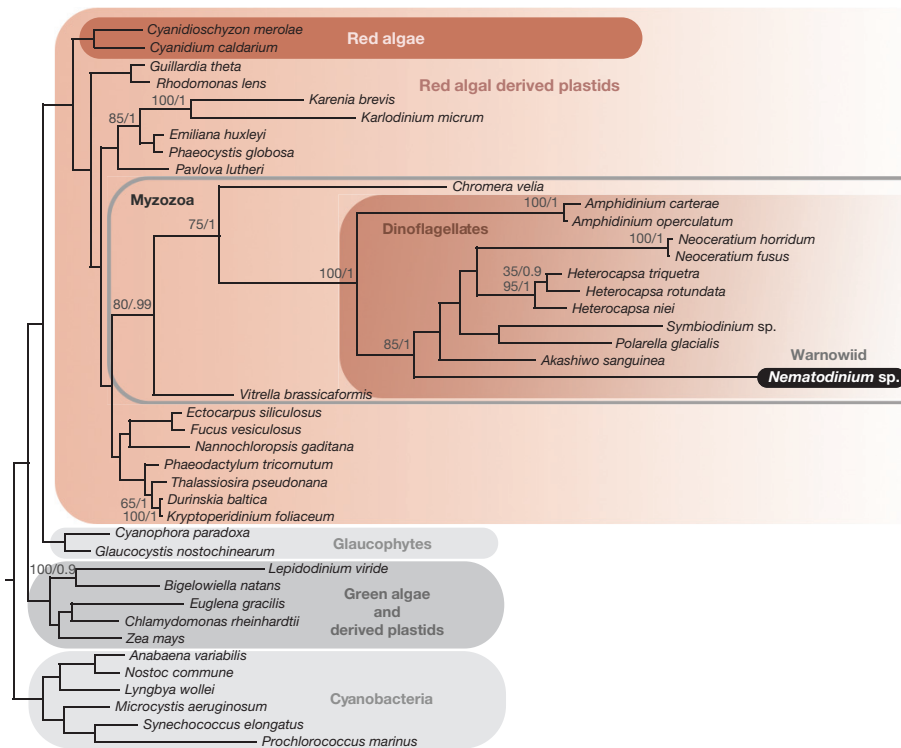
Figure 3 | Phylogeny of retinal-body-encoded proteins. Six partial plastid genes from the retinal body of the ocelloid in *Nematodinium* sp. were amplified. Photosystem I P700 apoprotein A2, photosystem II protein D1, photosystem II CP47 protein, photosystem II protein D1, cytochrome b_6, and cytochrome b_6/f complex subunit 4 were translated and concatenated for a 1,618-amino-acid alignment. The tree was inferred by analysing the 42-taxon alignment using maximum likelihood. Statistical support for the branches was evaluated using 500 maximum likelihood bootstrap replicates and Bayesian posterior probabilities. Support values are shown for all branches within the Myzozoa (dinoflagellates and chromerids).

Tomographic reconstructions also confirmed a close association between mitochondria and the lens of the ocelloid. The mitochondria surrounding the lens were interconnected and formed a sheet-like 'cornea' layer consistent with TEM data. The corneal layer surrounded all regions of the lens except for a few minor perforations and the side facing the retinal body (Fig. 4). The corneal mitochondria appear to form a continuous network with mitochondria in the nearby cytoplasm. The ocelloid, therefore, represents an intriguing mixture of components with endogenous and endosymbiotic origins.

Before this study, there was little evidence for homology between the ocelloid and other structures found in dinoflagellates[4]. On the basis of its resemblance to camera-type eyes, a relationship was even suggested between the ocelloid and the eyes of some animals[10]. To the contrary, our findings indicate that the ocelloid is a conglomerate of several membrane-bound organelles, including endomembrane vesicles, mitochondria, and plastids. The ocelloid is probably homologous to the much simpler eyespots found in several other lineages of dinoflagellates (Extended Data Fig. 9), most of which share features in
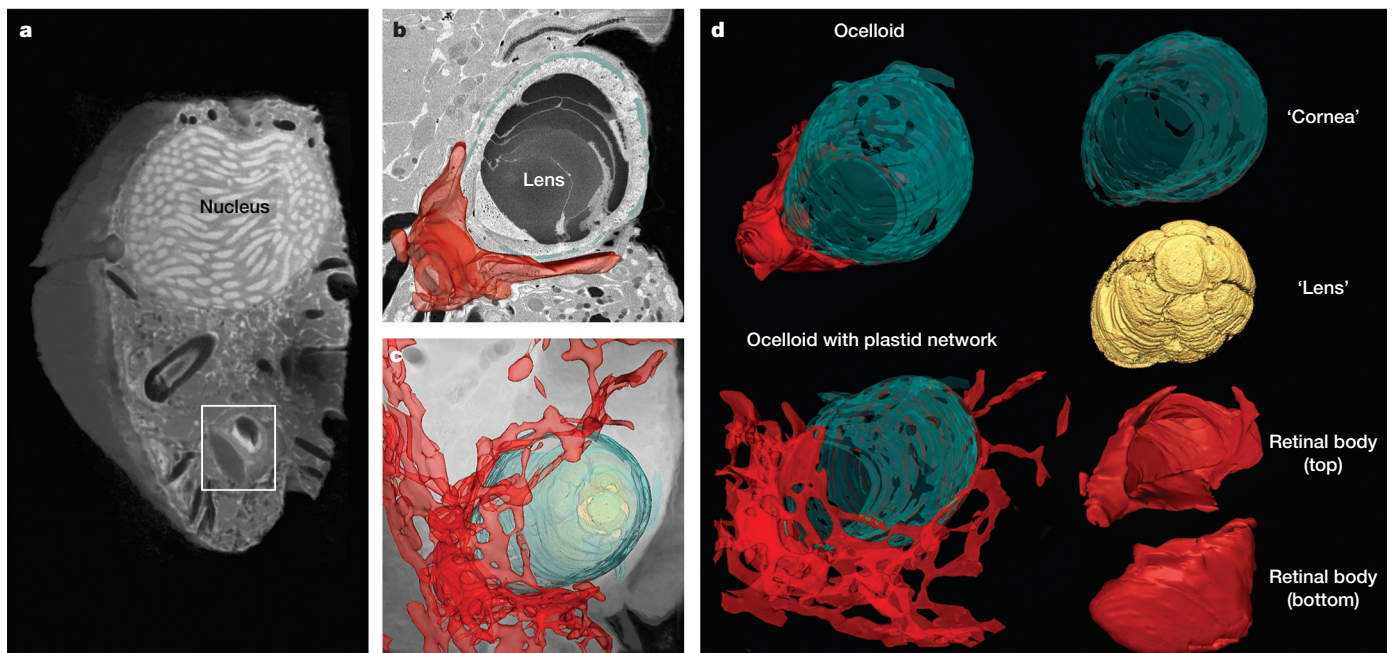
Figure 4 | Three-dimensional reconstruction of the ocelloid of *Nematodinium* sp. using FIB-SEM tomography. a, Stack of a halved cell, showing the nucleus and the ocelloid (box). **b**, FIB-SEM slice of the ocelloid, depicting the lens, mitochondria (blue), and retinal body (red). **c**, Translucent FIB-SEM stack of the region surrounding the ocelloid, including the lens (yellow) and full plastid network (red). **d**, Reconstructions of the ocelloid and its component parts, including the mitochondrial cornea-like layer, vesicular lens, and retinal body.

common with the peridinin plastid[4,19,20]. Peridinin plastids stem from an ancient red alga that was incorporated by the common ancestor of all myzozoans (dinoflagellates, chromerids, and apicomplexans), many of which (including all apicomplexans) subsequently lost photosynthesis and reduced their plastids to cryptic, morphologically simple structures[2,16]. While morphological reduction is a common trend among endosymbiotic organelles, the ocelloid in warnowiids demonstrates that increased complexity can also arise.

To understand the function of the ocelloid, a basic knowledge of the life history of warnowiid dinoflagellates is required. Understanding warnowiid behaviour is a difficult problem, however, because their cells are rarely encountered, have never been cultivated, and degrade rapidly when removed from the plankton[11]. Nevertheless, we observed one important detail of warnowiid life history using TEM of individual cells isolated directly from the ocean. We found that the food vacuoles in *Nematodinium* contained trichocysts (Extended Data Fig. 10), which are defensive extrusive organelles found in dinoflagellates[21]. These data suggest that *Nematodinium* feeds on other dinoflagellates, so one hypothesis is that the ocelloid is involved in the detection of other dinoflagellates as prey. Some dinoflagellates are capable of bioluminescence[22], which may be what ocelloids detect, but all dinoflagellates contain a distinctively large nucleus of permanently condensed chromosomes, and these chromosomes polarize light[23]. An intriguing possibility is that the ocelloid can detect polarized light, and, by extension, preferred prey. Testing such a specific phototactic behaviour will be challenging until warnowiids are brought into culture. Nevertheless, the genomic and detailed ultrastructural data presented here have resolved the basic components of the ocelloid and their origins, and demonstrate how evolutionary plasticity of mitochondria and plastids can generate an extreme level of subcellular complexity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Greuet, C. Organisation ultrastructurale de l'ocelle de deux Peridiniens Warnowiidae, *Erythropsis pavillardi* Kofoid et Swezy et *Warnowia pulchra* Schiller. *Protistologica* **4,** 209–230 (1968).
2. Janouskovec, J. *et al.* A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl Acad. Sci. USA* **107,** 10949–10954 (2010).
3. Kofoid, C. A. & Swezy, O. The free-living, unarmoured dinoflagellates. *Mem. Univ. Calif.* **5,** 1–562 (1921).
4. Dodge, J. D. The functional and phylogenetic significance of dinoflagellate eyespots. *Biosystems* **16,** 259–267 (1984).
5. Kreimer, G. Reflective properties of different eyespot types in dinoflagellates. *Protist* **150,** 311–323 (1999).
6. Greuet, C. Structural and ultrastructural evolution of ocelloid of *Erythropsidinium-pavillardi*, Kofoid-and-Swezy (dinoflagellate Warnowiidae, Lindemann) during division and palintomic divisions. *Protistologica* **13,** 127–143 (1977).
7. Greuet, C. Structure fine de locelle d'*Erythropsis pavillardi* Hertwig, peridinien Warnowiidae Lindemann. *C.R. Acad. Sci.* **261,** 1904–1907 (1965).
8. Hoppenrath, M. *et al.* Molecular phylogeny of ocelloid-bearing dinoflagellates (Warnowiaceae) as inferred from SSU and LSU rDNA sequences. *BMC Evol. Biol.* **9,** 116 (2009).
9. Leander, B. S. Different modes of convergent evolution reflect phylogenetic distances. *Trends Ecol. Evol.* **23,** 481–482 (2008).
10. Gehring, W. J. New perspectives on eye development and the evolution of eyes and photoreceptors. *J. Hered.* **96,** 171–184 (2005).
11. Gomez, F., Lopez-Garcia, P. & Moreira, D. Molecular phylogeny of the ocelloid-bearing dinoflagellates *Erythropsidinium* and *Warnowia* (Warnowiaceae, Dinophyceae). *J. Eukaryot. Microbiol.* **56,** 440–445 (2009).
12. Yoon, H. S. *et al.* Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332,** 714–717 (2011).
13. Lasken, R. S. Genomic sequencing of uncultured microorganisms from single cells. *Nature Rev. Microbiol.* **10,** 631–640 (2012).
14. Kolisko, M. *et al.* Single-cell transcriptomics for microbial eukaryotes. *Curr. Biol.* **24,** R1081–R1082 (2014).
15. Hofmann, E. *et al.* Structural basis of light harvesting by carotenoids: peridinin-chlorophyll-protein from *Amphidinium carterae*. *Science* **272,** 1788–1791 (1996).
16. Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64,** 583–607 (2013).
17. Saldarriaga, J. F. *et al.* Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J. Mol. Evol.* **53,** 204–213 (2001).
18. Mornin, L. & Francis, D. Fine structure of *Nematodinium armatum*, a naked dinoflagellate. *J. Microsc.* **6,** 759–772 (1967).
19. Lindberg, K., Moestrup, O. & Daugbjerg, N. Studies on woloszynskioid dinoflagellates I: *Woloszynskia coronata* re-examined using light and electron microscopy and partial LSU rDNA sequences, with description of *Tovellia* gen. nov. and *Jadwigia* gen. nov. (Tovelliaceae fam. nov.). *Phycologia* **44,** 416–440 (2005).
20. Moestrup, O., Hansen, G. & Daugbjerg, N. Studies on woloszynskioid dinoflagellates III: on the ultrastructure and phylogeny of *Borghiella dodgei* gen. et sp nov., a cold-water species from Lake Tovel, N. Italy, and on *B. tenuissima* comb. nov. (syn. *Woloszynskia tenuissima*). *Phycologia* **47,** 54–78 (2008).
21. Hausmann, K. Extrusive organelles in protists. *Int. Rev. Cytol.* **52,** 197–276 (1978).
22. Abrahams, M. V. & Townsend, L. D. Bioluminescence in dinoflagellates: a test of the burglar alarm hypothesis. *Ecology* **74,** 258–260 (1993).
23. Liu, J. & Kattawar, G. W. Detection of dinoflagellates by the light scattering properties of the chiral structure of their chromosomes. *J. Quant. Spectrosc. Radiat. Transf.* **131,** 24–33 (2013).

Supplementary Information is available in the online version of the paper.

Author Contributions G.S.G., S.H., P.J.K. and B.S.L. designed the experiments. G.S.G. performed light microscopy, TEM, FIB-SEM, dissected-organelle and single-cell genomics, and phylogenetic analyses on specimens he collected in Canada, with resources and funding from B.S.L. and P.J.K. S.H. performed light microscopy, TEM, and transcriptomics on specimens she collected in Japan with resources and funding from T.G., and was supported in Canada by P.J.K. and B.S.L. R.A.W. prepared genomic libraries for sequencing and participated in single-cell genomics with funding from C.A.S. G.S.G. and B.S.L. wrote the manuscript and all authors participated in the drafting process.

METHODS

Collection. From 2005 to 2009, *Erythropsidinium* sp. and *Warnowia* sp. were collected from the marine water column in Suruga Bay (Numaza, Shizuoka), Japan. On an inverted light microscope, cells of *Erythropsidinium* sp. were identified on the basis of the presence of an ocelloid and a piston organelle (Extended Data Fig. 2b and Supplementary Video 1). Cells of *Warnowia* sp. were recognized as ocelloid-bearing cells encircled three or more times by a helical groove (Extended Data Fig. 4a and Supplementary Video 2). cDNA libraries from four cells of *Warnowia* sp. and two cells of *Erythropsidinium* sp. were prepared as described[24]. In the summer of 2012 and 2013, *Nematodinium* sp. was collected from surface water in Bamfield Inlet, Bamfield, British Columbia, Canada, with a 20 μm plankton net. Cells of *Nematodinium* sp. were identified on the basis of the presence of an ocelloid and nematocysts (Extended Data Fig. 4c). Uncultivated *Nematodinium* sp. cells containing putative prey organisms (visible as pigmented vacuoles) were chosen for TEM, so that their feeding habits could be inferred from intracellular remnants (Extended Data Fig. 10). In total, 12 cells of *Nematodinium* sp. were fixed and mounted individually for TEM, and 58 cells of *Erythropsidinium* sp. were obtained and mounted for TEM in groups.

Fluorescence and differential interference contrast microscopy. Red epifluorescence of the *Nematodinium* sp. retinal body was excited with a 505 nm argon laser on a Zeiss Axioplan inverted microscope (Extended Data Fig. 4a). Differential interference contrast observations of *Nematodinium* sp., *Warnowia* sp., and *Erythropsidinium* sp. were performed using the same microscope (Extended Data Fig. 4).

Single-cell TEM of uncultivated *Nematodinium* sp. Each isolated cell of *Nematodinium* sp. was micropipetted onto a slide coated with poly-L-lysine. Cells were fixed with 2% glutaraldehyde in filtered seawater for 30 min on ice. After two washes in filtered seawater, cells were post-fixed in 1% OsO_4 for 30 min. Cells were dehydrated through a graded series of ethanol (50%, 70%, 85%, 90%, 95%, 100%, 100%) at 10 min each, and infiltrated with a 1:1 acetone-resin mixture for 10 min. Cells were steeped in Epon 812 resin for 12 h, after which the resin was polymerized at 60 °C for 24 h to produce a resin-embedded cell affixed to the glass slide. Using a power drill, resin was shaved to a 1 mm³ block, which was removed from the glass slide with a fine razor. The block, containing a single cell, was superglued to a resin stub in the desired orientation for sectioning. Thin (45 nm) sections were produced with a diamond knife, post-stained with uranyl acetate and lead citrate, and viewed under a Hitachi H7600 TEM.

Isolation of the retinal bodies of *Nematodinium* sp. In preparation for single-organelle genomics, five cells of *Nematodinium* sp. with no visible prey contents were selected to minimize the chances of genetic contamination. Each cell of *Nematodinium* was micropipetted onto a slide in a droplet of TE buffer and affixed to a patch of poly-L-lysine. Cells were lysed with nuclease-free water. The nucleus and other cell contents were gently dislodged with rinses of TE buffer, leaving the retinal body behind for manual isolation (Fig. 1d). Unlike the retinal body, which is darkly pigmented, the cornea and mitochondria of the ocelloid are much smaller, transparent, and could not be isolated after cell lysis or tracked through rinse steps. Five different retinal bodies were isolated and pooled onto a new, sterile slide, and washed three times with TE buffer to remove as many other cellular remnants as possible.

Single-organelle genomics of *Nematodinium* sp. To test for the presence of a plastid genome in the retinal body, we performed a genomic amplification using phiX 29 polymerase (Repli-G mini kit, Qiagen) on five individually isolated retinal bodies that were then pooled together. We performed a control reaction by amplifying a pool of five whole cells of *Nematodinium* sp. using the same procedures as for the retinal bodies. The whole-cell amplification provided a measure of overall plastid DNA concentration, against which the retinal body plastid DNA concentration could be compared. To minimize amplification bias, each reaction was divided into four aliquots, run in parallel, and pooled after the 15 h amplification period. Paired end sequencing on an Illumina MiSeq yielded 9,798 reads from the retinal bodies, versus 501,338 reads from whole cells. From these reads, plastid genes were assembled using the *de novo* assembly program Ray[25], which fragmented the reads into a variety of hash sizes ('kmers'), then assembled them. We found the assembly from 53 base pair (bp) kmers to be optimal, recovering six partial plastid genes (Fig. 1d, e). To estimate the concentration of plastid reads in the

whole cell versus isolated retinal body amplifications, we counted plastid reads in Bowtie[26], a read mapping program, then divided them by the total number of reads sequenced from that reaction (Fig. 1d, e).

Molecular phylogenetic analyses. The six plastid genes, photosystem I P700 apoprotein A2 (*PsaB*), photosystem II protein D1 (*PsbA*), photosystem II CP47 protein (*PsbB*), photosystem II protein D1 (*PsbD*), cytochrome b_6 (*PetB*), and cytochrome b_6/f complex subunit 4 (*PetD*) were translated, and their amino acids aligned with a representative set of eukaryotes in Muscle[27], with fast-evolving and ambiguously aligned regions removed in Gblocks 0.91b[28]. GenBank accession numbers are listed in Extended Data Figs 6 and 7. The amino-acid substitution model (Protein GTR gamma) was estimated from the concatenated alignment of 1,618 amino acids using the Models package in Mega 6.0.5 (ref. 29). A maximum likelihood phylogeny was run with 500 bootstraps in RAxML[30]. A second, Bayesian analysis was run for 10,000 generations in MrBayes 3.2 (ref. 31), using the high-heating setting of (nchains = 4), to account for rapid evolution of dinoflagellate plastids. These maximum likelihood analyses were run both for the multiprotein data set and for each protein individually (Extended Data Figs 6 and 7). A dinoflagellate phylogeny was estimated using 18S and 28S ribosomal DNA sequences, concatenated as 2,331 nucleotide alignment, across 36 dinoflagellate taxa including published sequences from *Nematodinium* sp., *Warnowia* sp., and *Erythropsidinium* sp. (Extended Data Fig. 6).

FIB-SEM. Cells of *Nematodinium* sp. were individually transferred into a droplet of 20% bovine serum albumin in phosphate buffered saline solution (an osmotically inert solution). Cells were frozen immediately to minimize fixation artefacts, using a Leica EM HPM 100 high-pressure freezer. Freeze substitution was subsequently used to remove the aqueous content of the cells and replace it with an acetone solution containing 5% water, 1% osmium tetroxide, and 0.1% uranyl acetate, at −80 °C for 48 h, −20 °C for 6 h, then graded back to 4 °C over 13 h. The prepared samples were washed twice in 100% acetone. Two cells were recovered by micropipette. Each cell was placed on a separate Thermonox coverslip, where it adhered to a patch of poly-L-lysine. In preparation for FIB-SEM, cells were infiltrated with a 1:1 mix of acetone and Embed 812 resin for 2 h, then 100% resin overnight. A second Thermonox coverslip was applied, sandwiching each cell in a thin layer of resin between the coverslips. Resin was polymerized at 65 °C for 24 h. The top coverslip was then removed with a razor blade to expose the resin face overlying the cell.

A single cell was imaged by an FEI Helios NanoLab 650 dual-beam FIB-SEM. The ion beam milled through the cell in 20 nm increments, yielding 190 image slices. Slices were aligned as a *z*-stack in Amira 5.5. Features of interest, including mitochondria and chloroplasts, were semi-automatically segmented: that is, manually traced in approximately one of every three slices, before automatic interpolation filled in the volumes between the slices. Images that did not pass quality screening because of fluctuations in microscope beam power and autofocus were not directly segmented, but were interpolated from segmentation on neighbouring images, according to the manufacturer's instructions. Surfaces of the mitochondria, chloroplasts, and vesicles were generated, smoothed, and colourized to produce a three-dimensional model of the components that form the ocelloid (Supplementary Video 3).

24. Hayakawa, S. *et al.* Function and evolutionary origin of unicellular camera-type eye structure. *PLoS ONE* **10,** http://dx.doi.org/10.1371/journal.pone.0118415 (2015).
25. Boisvert, S. *et al.* Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* **17,** 1519–1533 (2010).
26. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
27. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
28. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17,** 540–552 (2000).
29. Tamura, K. *et al.* MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).
30. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
31. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61,** 539–542 (2012).
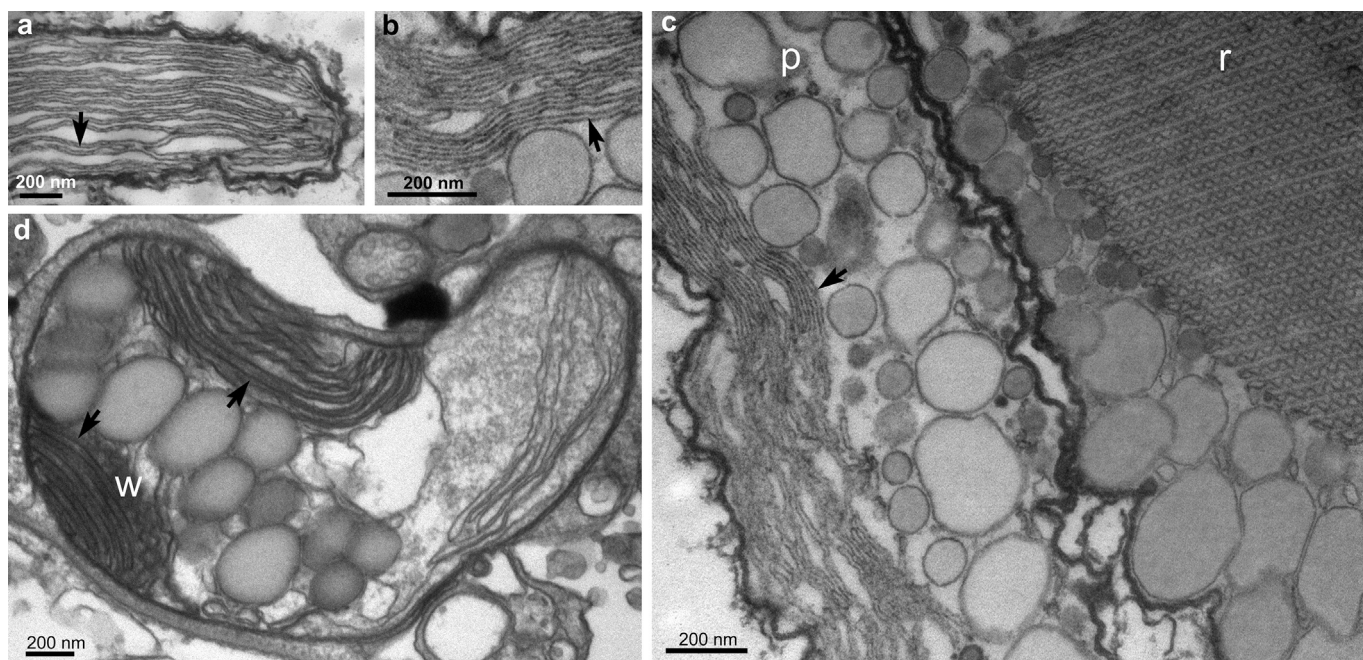
Extended Data Figure 1 | TEM of thylakoid membranes in *Nematodinium* sp. a, A small, peripheral plastid in *Nematodinium* sp. with typical thylakoids resembling peridinin plastids in other dinoflagellates. **b**, Thylakoids in the iris region of the ocelloid. **c**, Thylakoids in the iris positioned beside waveform membranes (w) of the retinal body, during interphase. **d**, A retinal body towards the end of interphase, in which the waveform membranes de-differentiate and are continuous with the typical thylakoids. Typical thylakoids are marked by arrows.

a *Nematodinium* sp.
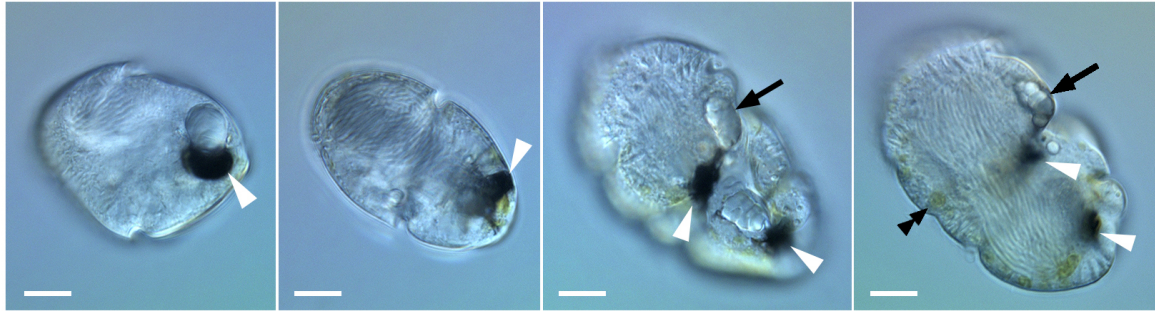
INTERPHASE (differentiated)   DIVISION (dedifferentiated)

b *Erythropsidinium* sp.
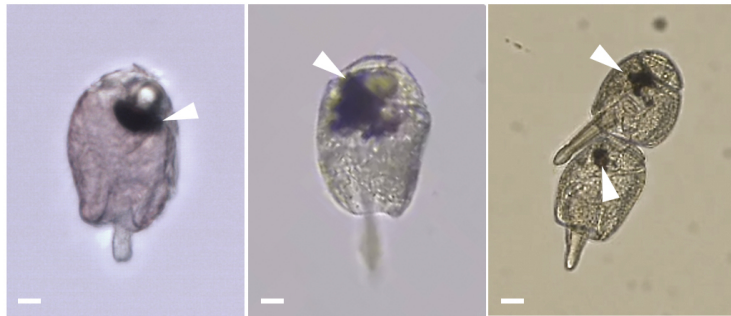
c

waveform membranes   paired, vertical thylakoids   normal thylakoids
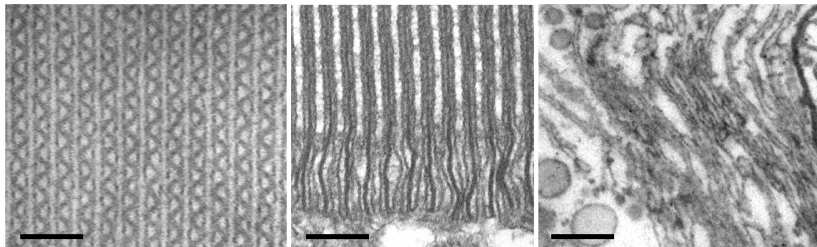
Extended Data Figure 2 | Development in warnowiids a, b, Light micrographs of several cells of *Nematodinium* sp., and *Erythropsidinium* sp., progressing from interphase (left) to division (right). Scale bars, 10 μm. **c,** TEM of membranes in the retinal body, during differentiated (left, *Nematodinium* sp.), transitional (middle, *Erythropsidinium* sp.), and de-differentiated modes (right, *Nematodinium* sp.). Scale bars, 200 nm. The double arrowhead marks a typical plastid; arrowheads mark the retinal bodies; arrows mark lenses that are de-differentiating.
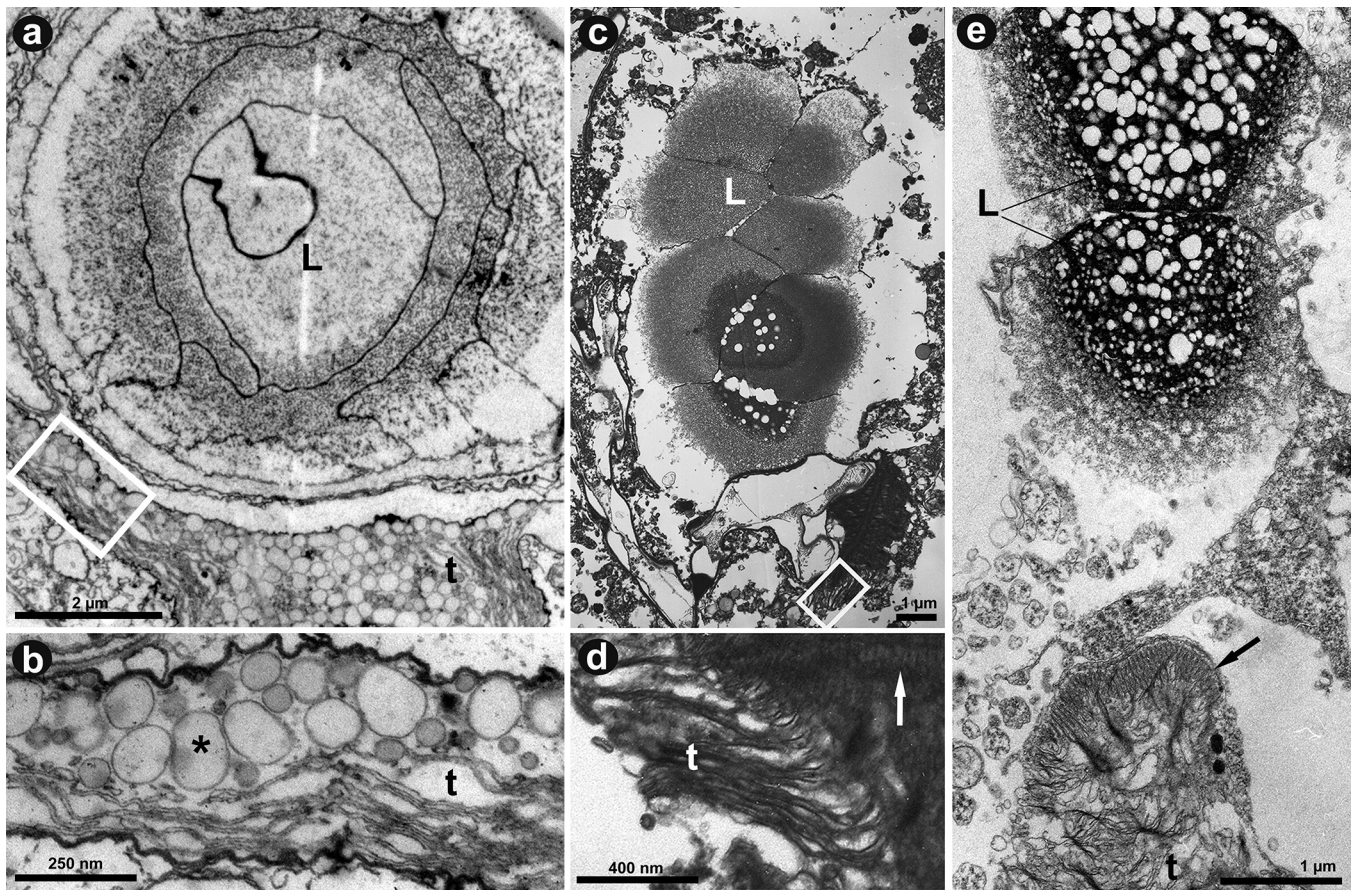
Extended Data Figure 3 | Transient thylakoids in the retinal body viewed with TEM. a, b, Ocelloid in a cell of *Nematodinium* sp. near division. **c–e**, Ocelloid in cells of *Erythropsidinium* sp. during division. L, lens; t, thylakoids; asterisks, lipid droplets; arrows, waveform membranes.
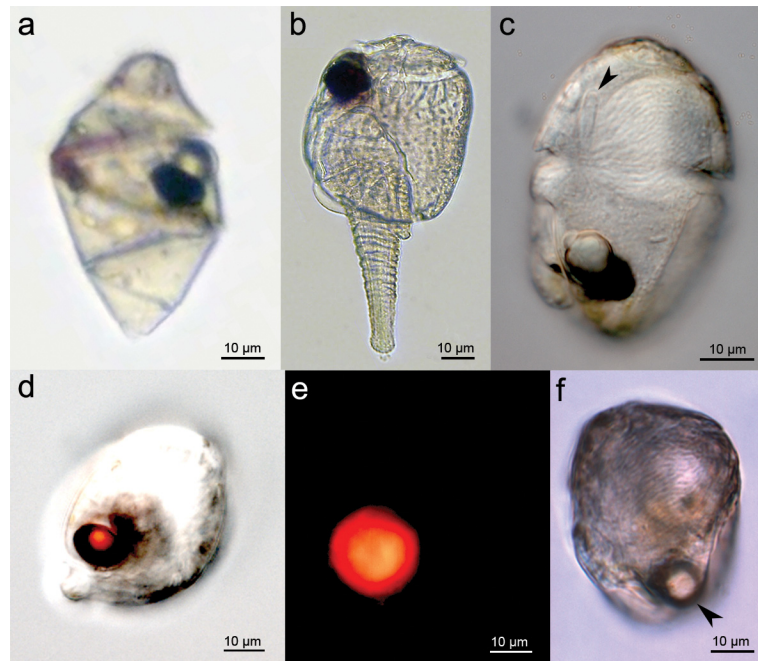
Extended Data Figure 4 | Light micrographs of warnowiids used in this study. a, Still frame from a video of *Warnowia* sp. **b**, *Erythropsidinium* sp. **c**, *Nematodinium* sp. with a nematocyst (arrowhead). **d**, The ventral side of *Nematodinium* sp. showing red pigmentation of the retinal body. **e**, Epifluorescence image of the same cell and angle, showing red fluorescence of the retinal body excited by 505 nm light. **f**, *Nematodinium* sp. showing a bright spot of reflectivity (that is, 'eyeshine') (arrowhead) in the ocelloid. Scale bars, 10 μm.
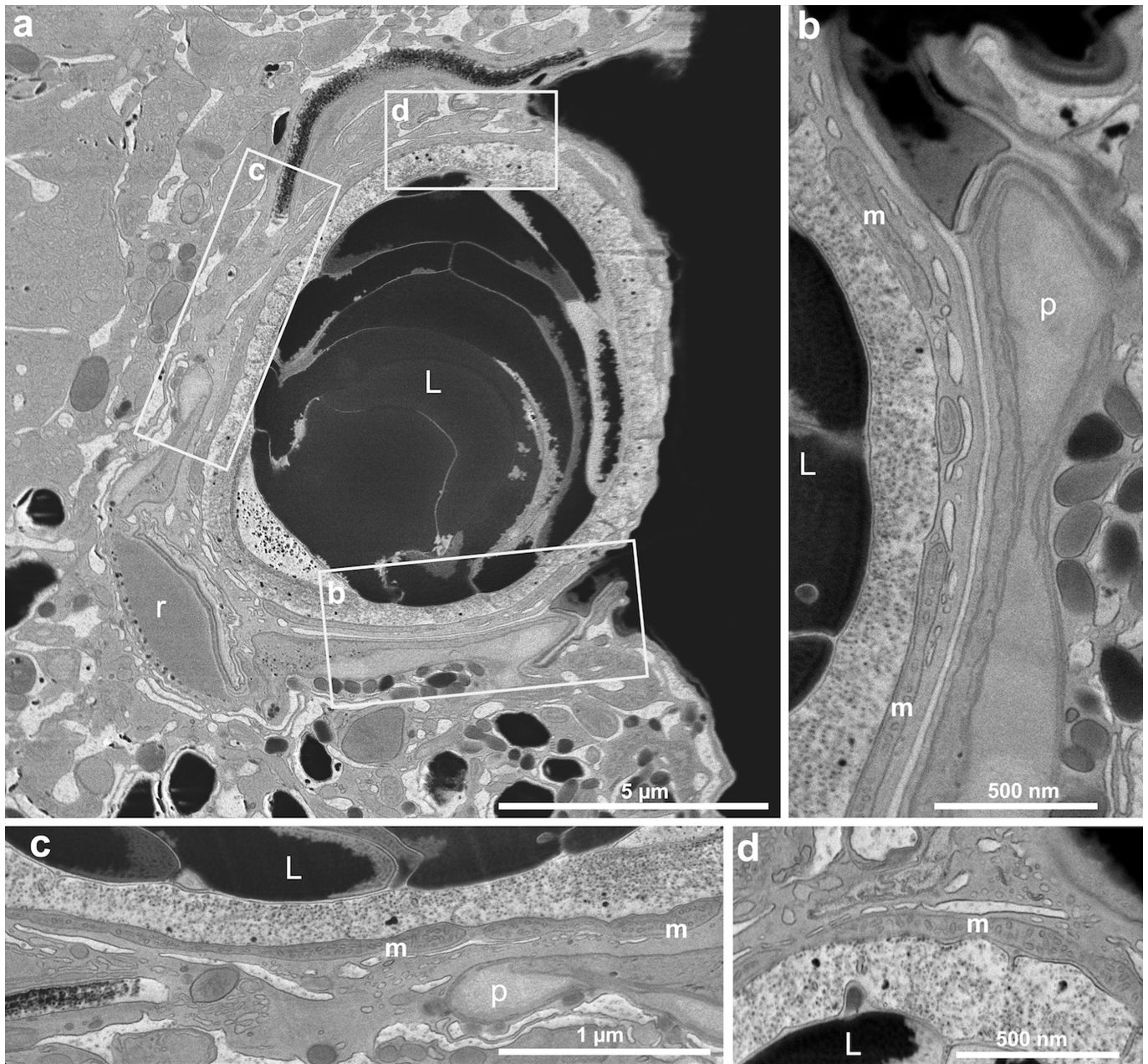
Extended Data Figure 5 | TEM of the cornea-like layer of mitochondria in the ocelloid of *Nematodinium* sp. **a**, Low-magnification TEM of the ocelloid, with rectangles delimiting the areas of higher magnification shown in **b–d**. **b–d**, High magnifications of structures bordering the lens (L). Mitochondria, m; pigmented ring, p; retinal body, r.
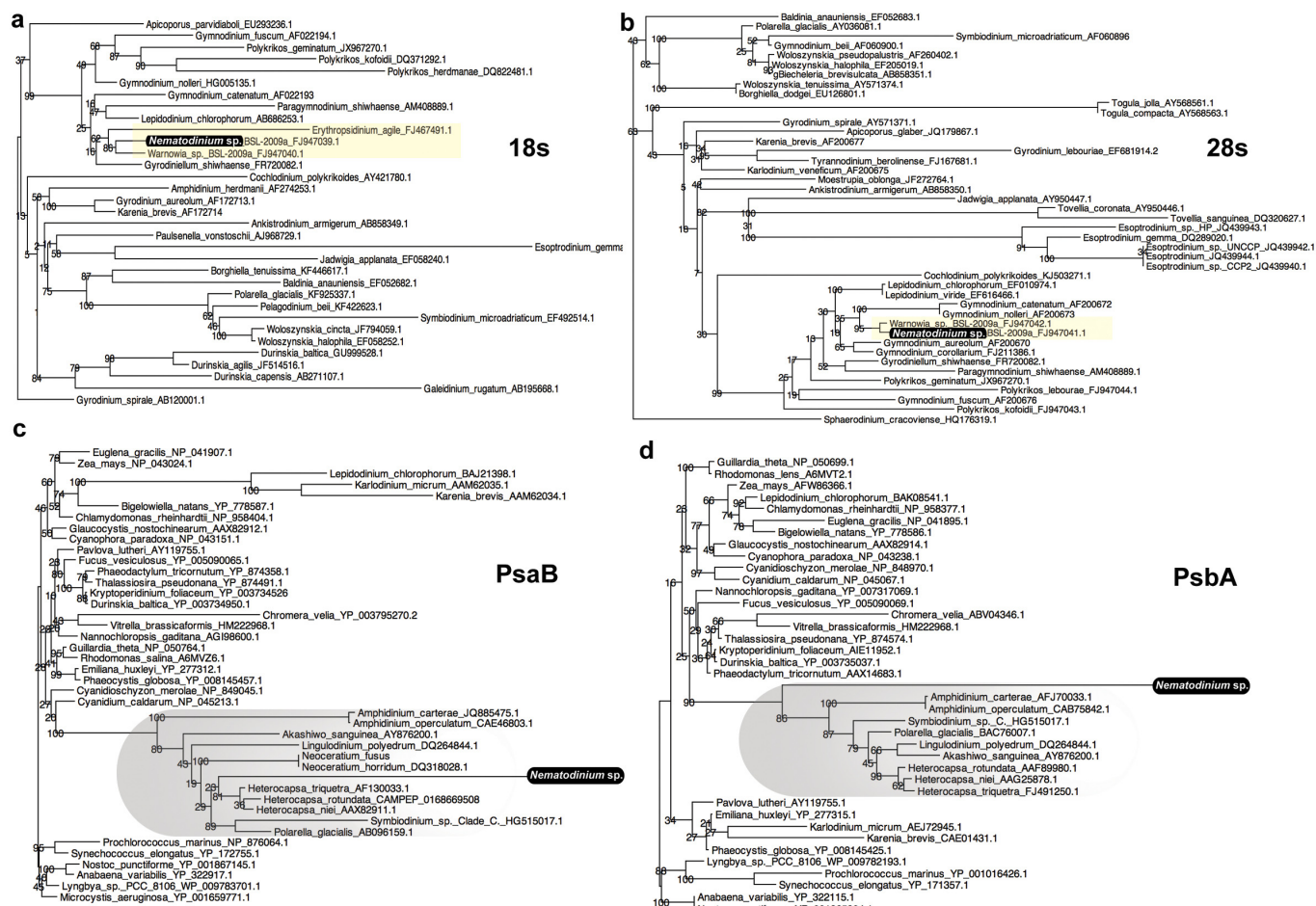
Extended Data Figure 6 | Individual ribosomal gene and photosystem protein gene trees. For **c** and **d**, the photosystem genes for *Nematodinium* sp. were amplified from the retinal body of the ocelloid. Support values for all phylogenies were calculated from 100 bootstraps using maximum likelihood analysis. **a**, 18S ribosomal DNA gene phylogeny derived from a 1,717-bp alignment across 33 dinoflagellate taxa. **b**, 28S ribosomal DNA gene phylogeny derived from a 970-bp alignment across 43 dinoflagellate taxa. For both **a** and **b**, warnowiids are highlighted in yellow and *Nematodinium* sp. is highlighted in black. **c**, Photosystem I P700 apoprotein A2 (*PsaB*) protein phylogeny derived from a 508 amino acid (AA) alignment across 42 photosynthetic taxa. **d**, Photosystem II protein D1 (*PsbA*) protein phylogeny derived from a 360 AA alignment across 39 photosynthetic taxa. For **c** and **d**, dinoflagellates are shaded in grey, and *Nematodinium* sp. is highlighted in black.
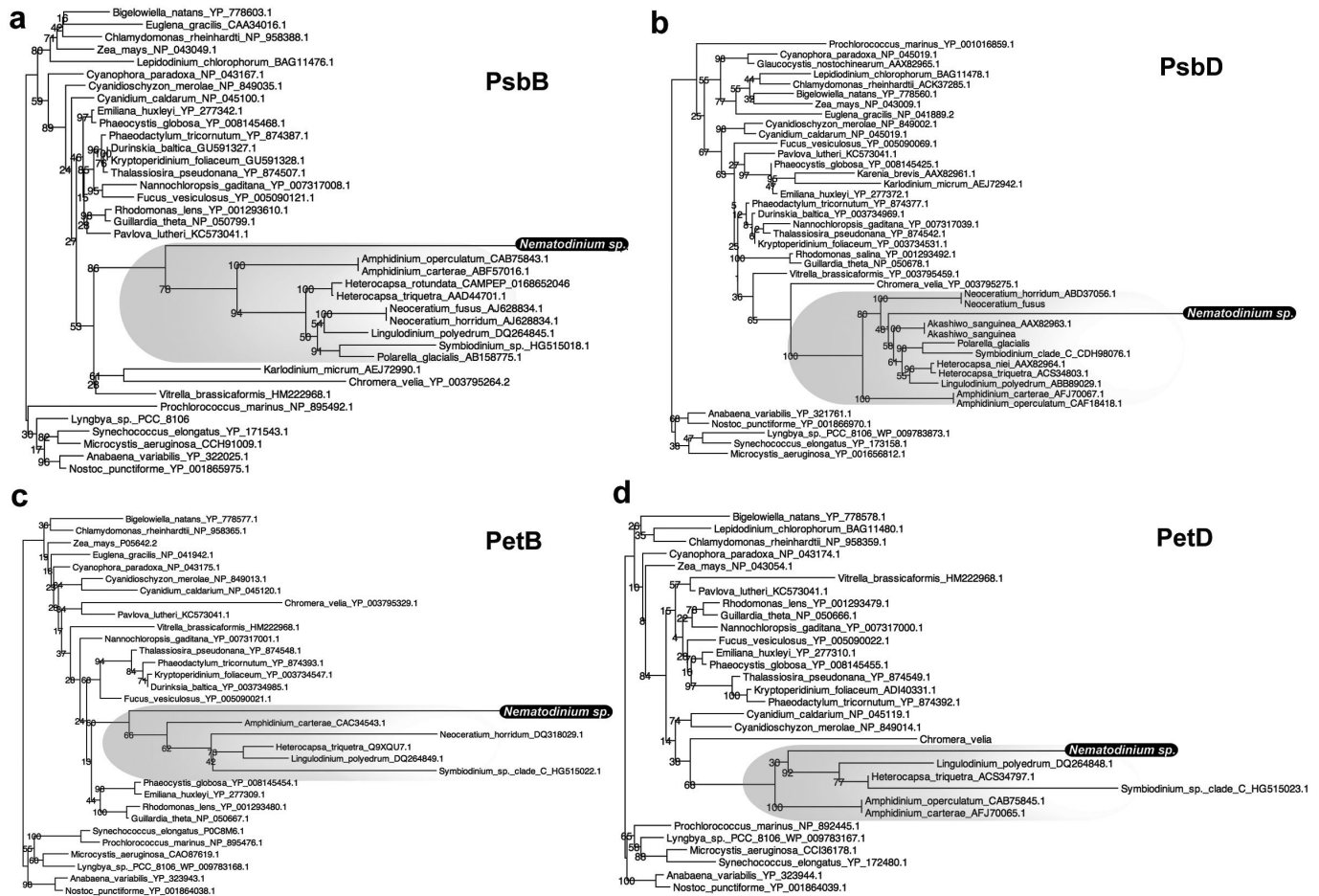
Extended Data Figure 7 | Individual photosystem protein trees. All the photosystem genes from *Nematodinium* sp. were amplified from the retinal body of the ocelloid. Support values for all phylogenies were calculated from 100 bootstraps using maximum likelihood analysis. **a**, Photosystem II CP47 (*PsbB*) protein phylogeny derived from a 504 AA alignment across 38 photosynthetic taxa. **b**, Photosystem II protein D1 (*PsbD*) phylogeny derived from a 342 AA alignment across 42 photosynthetic taxa. **c**, Cytochrome b_6 (*PetB*) protein phylogeny derived from a 216 AA alignment across 32 photosynthetic taxa. **d**, Cytochrome b_6/f complex subunit 4 (*PetD*) protein phylogeny derived from an 161 AA alignment across 31 photosynthetic taxa. Dinoflagellates are shaded in grey, and *Nematodinium* sp. is highlighted in black.
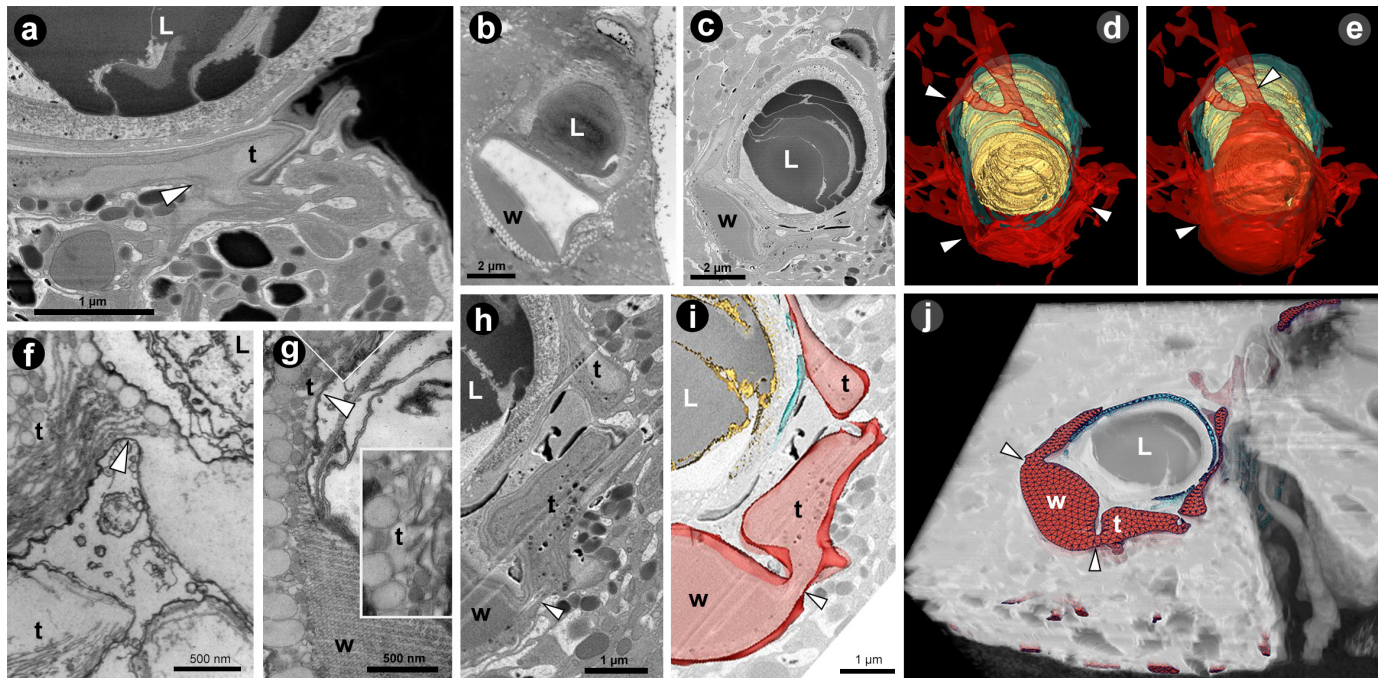
Extended Data Figure 8 | Continuity between the retinal body and the plastid network in *Nematodinium* sp. a, FIB-SEM slice of plastids attached to retinal body. **b**, TEM overview of ocelloid in a high-pressure frozen cell. **c**, FIB-SEM overview of ocelloid in a high-pressure frozen cell. **d**, Three-dimensional reconstruction of the ocelloid shown halved. **e**, Three-dimensional reconstruction of the ocelloid in full. **f**, Fusion site between plastids joined to the retinal body as seen in TEM. **g**, Site where the waveform-membrane region of the ocelloid joins to a region with thylakoids as seen in TEM. Inset shows thylakoids, and corresponds to the box in the main image. **h**, Fusion site as seen through FIB-SEM. **i**, Tracing of membrane continuity in Amira. **j**, Partial reconstruction of the ocelloid in Amira. Arrowheads point to fusion zones between sites bounded by the plastid membrane (reconstructed in red), blue denotes mitochondria, yellow denotes the surface of the lens. L, lens; w, waveform membranes; t, thylakoids.
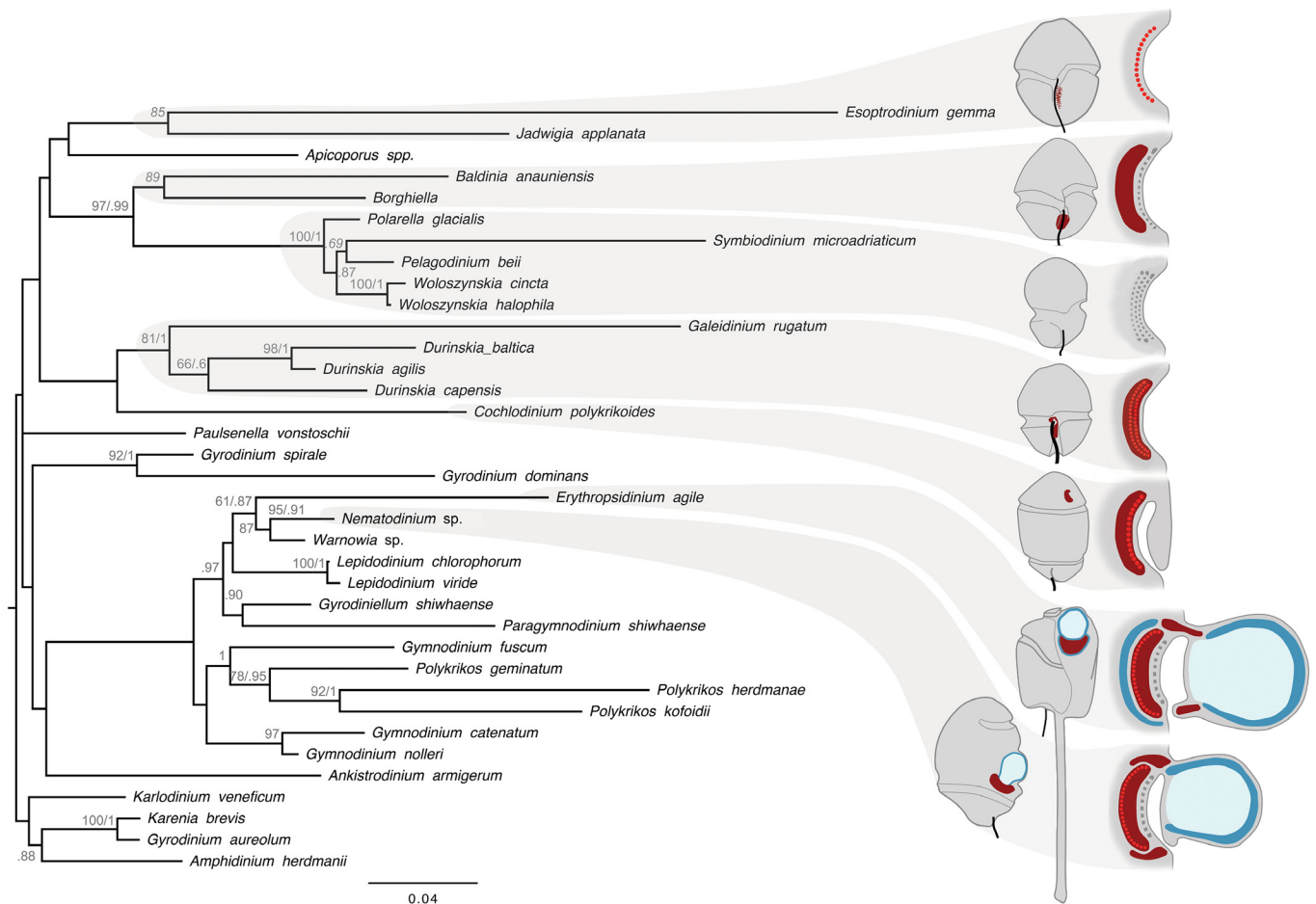
Extended Data Figure 9 | Dinoflagellate eyespot types within a phylogenetic context. Diagrams of whole cells and eyespots are shown for all dinoflagellates for which both ultrastructural descriptions and 18S and 28S ribosomal DNA sequences have been published. Eyespot diagrams highlight plastid-like structures (crimson), as well as mitochondria (dark blue), lens-like vesicles (light blue), lipid droplets (red dots), and crystalline layers (grey dashes). The phylogenetic tree was inferred from a 2,331-nucleotide alignment of concatenated 18S and 28S ribosomal DNA sequences across 36 genera; statistical support was evaluated with 500 bootstraps using maximum likelihood and 10,000 generations of Bayesian analysis. Bootstrap values above 60% are shown. For some taxa, 18S and 28S ribosomal sequences were concatenated from different species within the genus. Only the genus is shown for these taxa.
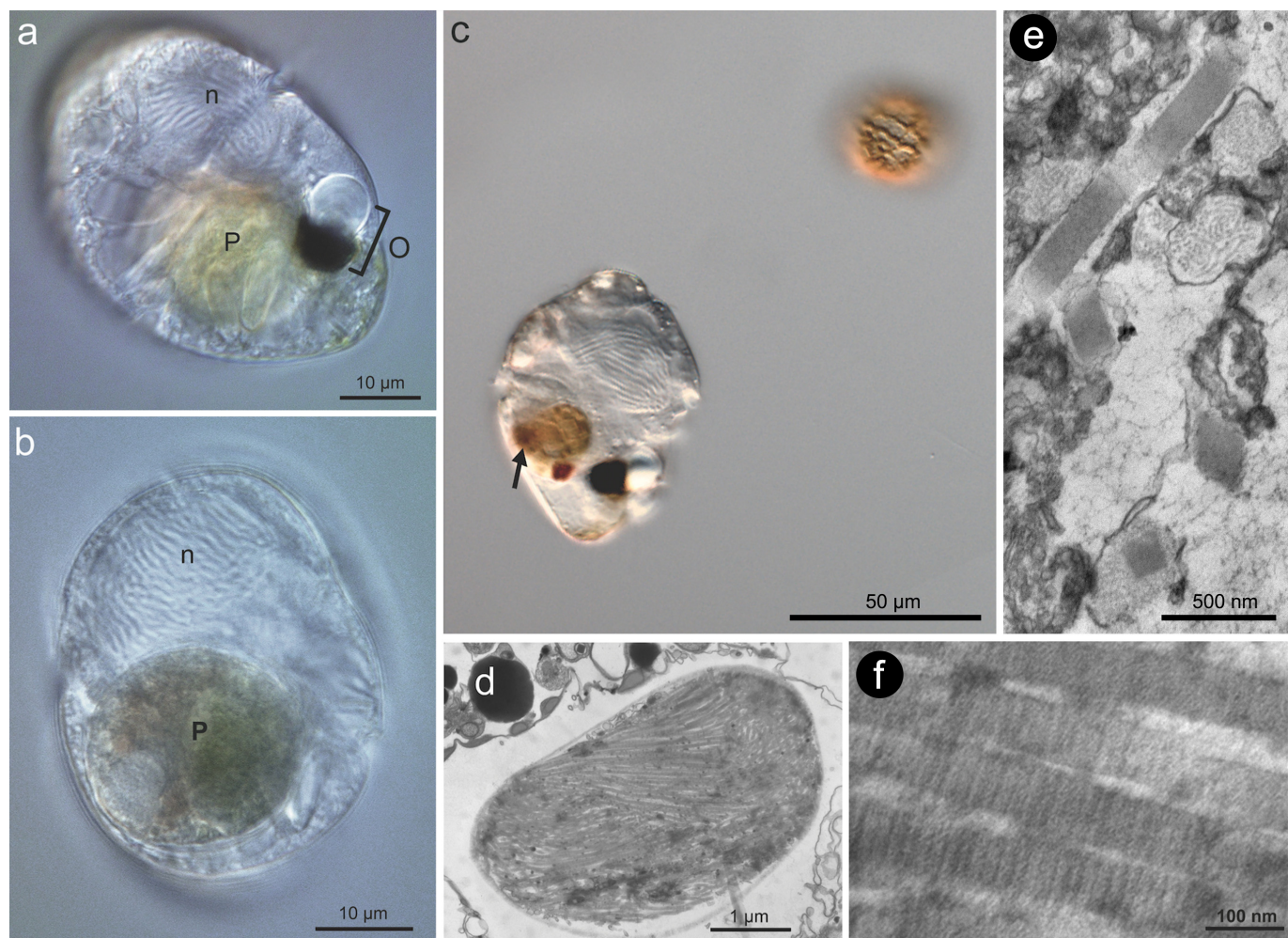
Extended Data Figure 10 | Light micrographs and TEM showing food vacuoles in *Nematodinium* sp. **a**, Differential interference contrast light micrographs showing a cell with prey (P) visible as green tinted food vacuole. **b**, Differential interference contrast light micrographs showing a cell in which the condensed dinoflagellate-type nuclei (n) are visible as birefringent chromosomes both in the predator and in the prey. **c**, Differential interference contrast light micrographs of a *Nematodinium* sp. cell containing digested prey (arrowhead) and co-occurring with potential prey, a smaller dinoflagellate. **d**, TEM showing a food vacuole inclusion consisting of a bolus of discharged trichocysts. **e**, TEM of undischarged dinoflagellate-type trichocysts showing their characteristic square shape in transverse section. **f**, TEM of discharged dinoflagellate-type trichocysts showing their characteristic striation pattern in longitudinal section.

LETTER

Unusual biology across a group comprising more than 15% of domain Bacteria

Christopher T. Brown[1], Laura A. Hug[2], Brian C. Thomas[2], Itai Sharon[2], Cindy J. Castelle[2], Andrea Singh[2], Michael J. Wilkins[3,4], Kelly C. Wrighton[4], Kenneth H. Williams[5] & Jillian F. Banfield[2,5,6]

A prominent feature of the bacterial domain is a radiation of major lineages that are defined as candidate phyla because they lack isolated representatives. Bacteria from these phyla occur in diverse environments[1] and are thought to mediate carbon and hydrogen cycles[2]. Genomic analyses of a few representatives suggested that metabolic limitations have prevented their cultivation[2–6]. Here we reconstructed 8 complete and 789 draft genomes from bacteria representing >35 phyla and documented features that consistently distinguish these organisms from other bacteria. We infer that this group, which may comprise >15% of the bacterial domain, has shared evolutionary history, and describe it as the candidate phyla radiation (CPR). All CPR genomes are small and most lack numerous biosynthetic pathways. Owing to divergent 16S ribosomal RNA (rRNA) gene sequences, 50–100% of organisms sampled from specific phyla would evade detection in typical cultivation-independent surveys. CPR organisms often have self-splicing introns and proteins encoded within their rRNA genes, a feature rarely reported in bacteria. Furthermore, they have unusual ribosome compositions. All are missing a ribosomal protein often absent in symbionts, and specific lineages are missing ribosomal proteins and biogenesis factors considered universal in bacteria. This implies different ribosome structures and biogenesis mechanisms, and underlines unusual biology across a large part of the bacterial domain.

We sampled microbial communities from an aquifer adjacent to the Colorado River near the town of Rifle, Colorado, USA in 2011. Groundwater was filtered through a 1.2 μm pre-filter and cells were collected on serial 0.2 and 0.1 μm filters (Extended Data Fig. 1). Post-0.2 μm filtrates were targeted because CPR bacteria were predicted to have ultra-small cells on the basis of their small genomes[2]. Groundwater was sampled before and during an acetate amendment experiment that reproduced conditions that generated the first genomes from CPR bacteria[2,4,7,8] (Supplementary Table 1). Total DNA and RNA were extracted from filters and sequenced. We obtained 224 gigabase pairs (Gb) of paired-end metagenomic sequence from 12 samples (150 bp reads, 6 time points, 0.2 and 0.1 μm filters; Supplementary Table 2). Sequence assembly generated 3.9 Gb of contiguous sequences ≥5 kb. We also obtained 181 Gb of metatranscriptomic sequence from six samples (50 bp reads, 0.2 μm filters).

Assembled scaffolds were binned into genomes on the basis of their GC content, DNA sequence coverage, abundance pattern across samples, and taxonomic affiliation (binning was validated with a tetranucleotide sequence signature method; Extended Data Fig. 2). Overall, we reconstructed >1,750 genome bins from microbial community sequence data. Here, we focus on genomes from CPR bacteria and TM6, which represented >60% of bins. Included in our analyses of the CPR are members of the Parcubacteria (OD1), Microgenomates (OP11), WWE3, Berkelbacteria (ACD58), Saccharibacteria (TM7), WS6, Peregrinibacteria (PER), and Kazan phyla, in addition to previously unrecognized lineages (CPR1–3; Fig. 1). In total, 789 draft-quality (≥50% complete) genomes were reconstructed (Table 1). We manually curated eight genomes to completion: the first three from Microgenomates, two from Parcubacteria, one each from Kazan and Berkelbacteria, and an additional genome from Saccharibacteria. All complete and draft genomes are small and most are <1 Mb in length (Supplementary Tables 3 and 4).

In total, 1,543 bacterial 16S rRNA genes ≥800 bp were assembled and curated to eliminate assembly errors (713 sequences clustered at 97% identity; Supplementary Data 1). Relative abundance measurements show enrichment of CPR organisms in small-cell filtrates, suggesting that they have ultra-small cells (Extended Data Fig. 3). This finding is supported by a recent microscopy study[8]. Surprisingly, 31% of 16S rRNA genes encoded a large (≥10 bp) insertion sequence (maximum 2,004 bp; mean 519 bp; standard deviation (s.d.) 372 bp; Supplementary Table 5). Insertions are found in phylogenetically diverse members of CPR phyla (Fig. 1, Supplementary Fig. 1 and Supplementary Data 2). Insertion sites are clustered in several distinct locations on the 16S rRNA gene, both in variable and conserved regions (Fig. 2). Most insertions ≥500 bp encode a catalytic RNA intron (group I or II) and/or an open reading frame (ORF), suggesting that they are self-splicing. Encoded proteins frequently belong to families of homing endonucleases (LAGLIDAG 1–3 and GIY-YIG). However, 25% are not similar to known protein families or to each other. These may represent novel endonucleases or may no longer be functional, since loss of function is common in homing endonucleases[9].

Four members of the Thiotrichaceae family are the only bacteria known to have self-splicing introns within their 16S rRNA genes[10]. An extensive search for insertions in genes from our study and the Silva database[11] suggests their rarity in bacteria outside the CPR (Extended Data Fig. 4 and Supplementary Table 6). Especially rare are insertions encoding predicted self-splicing introns and/or ORFs. However, these genes need not be functional if the genome encodes additional, insertion-free copies. Importantly, all complete CPR genomes have only one copy of the 16S rRNA gene (this study and others[3,6]). Sequencing coverage analysis of draft genomes further indicates that a single copy is typical for these lineages (Extended Data Fig. 5 and Supplementary Table 7).

Mapping metatranscriptomic sequences to assembled 16S rRNA genes showed that insertions are not retained in transcribed RNAs and are probably rapidly degraded (Supplementary Table 8). However, it is possible that spliced sequences are rendered inaccessible to sequencing after hybridizing, circularizing, or, in some cases, due to their small size. Regardless of their fate, splicing establishes these insertions as introns. Self-splicing is expected if insertions encode a catalytic RNA intron; however, splicing could also occur via an RNase-III-mediated mechanism[12]. Several genes contain multiple introns. For example, one of the complete genomes we obtained encodes a 16S rRNA gene with four introns (Fig. 3).

[1]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. [2]Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA. [3]School of Earth Sciences, The Ohio State University, Columbus, Ohio 43210, USA. [4]Department of Microbiology, The Ohio State University, Columbus, Ohio 43210, USA. [5]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [6]Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720, USA.
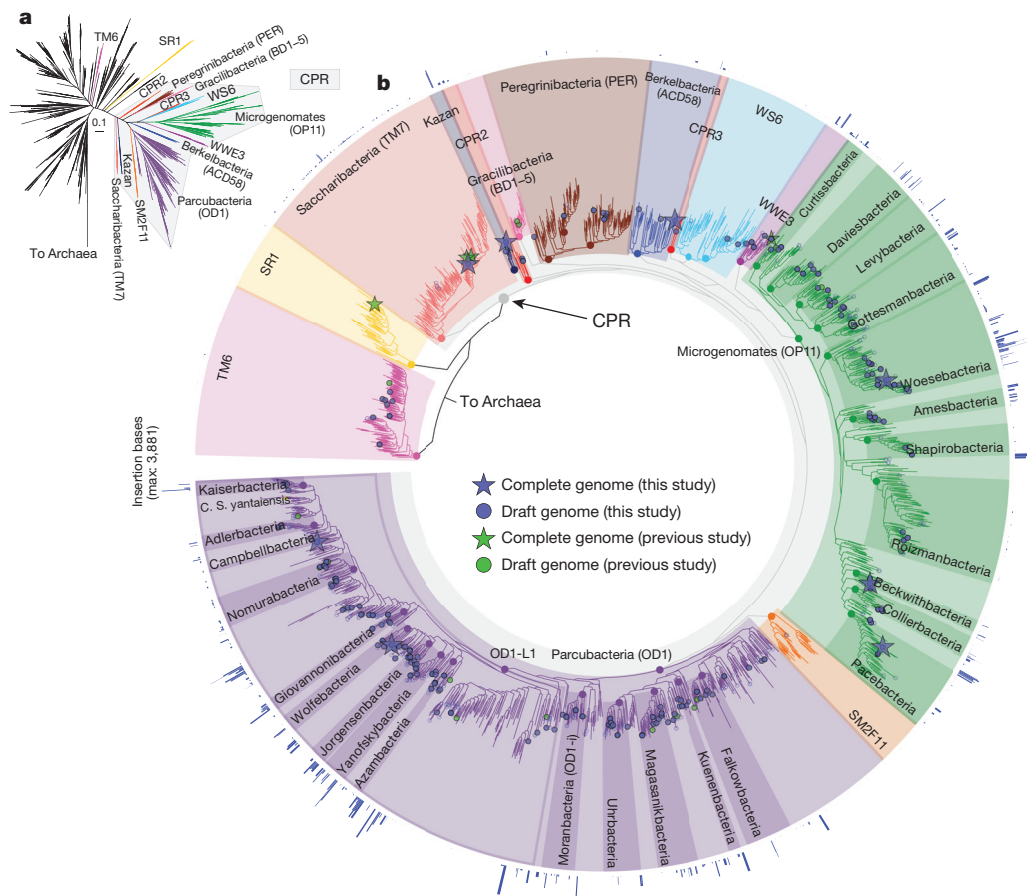
Figure 1 | Phylogeny and genomic sampling of the CPR. a, b, Subsets of a maximum-likelihood 16S rRNA gene phylogeny (Supplementary Fig. 1) showing the CPR, a monophyletic radiation of candidate phyla (**a**), and genomic sampling of candidate phyla (**b**). Proposed names for phyla within the superphyla Parcubacteria and Microgenomates are explained in Extended Data Table 1. Many CPR 16S rRNA genes encode insertions (length shown by blue bars, combined length for multiple insertions).

CPR bacteria frequently encode introns in 23S rRNA genes with features similar to those in 16S rRNA genes (Extended Data Fig. 6, Supplementary Tables 5, 8 and Supplementary Data 3). However,

Figure 2 | Features of insertions encoded within CPR 16S rRNA genes.
Insertions identified in assembled, unique bacterial 16S rRNA genes occur in conserved and variable (V; red bars) regions (Supplementary Table 5). Histograms show the frequency of insertions. Insertions are of several types distinguishable by catalytic RNA introns and/or ORFs. IVP, intervening sequence protein.

these introns and encoded proteins share little sequence similarity with one another (Supplementary Table 9). It remains a puzzle why introns in critical, highly transcribed rRNA genes do not make these organisms uncompetitive, as their transcription is costly, even though formation of nonfunctional ribosomes is avoided by splicing.

Insertions in rRNA genes are found in *Coxiella* and Rickettsiales-lineage endosymbionts[13,14]. Interestingly, one member of the Parcubacteria, '*Candidatus* Sonnebornia yantaiensis', is intracellular[15], but does not contain an insertion in its 16S rRNA gene (Fig. 1). However, there is no evidence that an intercellular lifestyle is typical across CPR lineages, although a strong dependence on other community members is likely[3,4].

Metagenomic analyses are polymerase chain reaction (PCR)-independent and, therefore, not biased by primers designed on the basis of expectations of sequence conservation. As a consequence, our sampling indicated that many CPR organisms would evade detection by 16S rRNA gene amplicon surveys. Primer binding analysis showed that primers extensively used in microbial surveys (515F and 806R[16]) would probably not bind to 16S rRNA genes of ~50% of Microgenomates, ~50% of Saccharibacteria, 60% of WWE3, and 100% of WS6 sequences sampled here (Extended Data Fig. 7). In fact, these primers would probably miss ~20% of all bacteria detected in this study, including organisms outside the CPR. Furthermore, introns in these genes would interfere with amplification, both because they occur in regions targeted by primers and because they increase the length of the target sequence. In addition to being excluded during size-selection of amplicons, intron-containing genes are less likely to amplify compared with shorter, intron-free genes[10]. Thus, several barriers have prevented identification of many CPR bacteria.

Removal of introns from 16S rRNA gene sequences, followed by structural alignment[17], was critical to establishing a reliable phylogeny. The new phylogenetic analysis shows that the CPR is monophyletic (Fig. 1), a result also evident in concatenated ribosomal protein

Figure 3 | Intron-encoding 16S rRNA gene from complete Microgenomates genome. a, Stringent mapping of paired-read metagenome sequences confirms the assembly. **b**, 16S rRNA encoding regions, but not insertions, are covered by perfectly matched metatranscriptome sequences. The absence of RNA sequences for insertions indicates that they are introns. Shown are regions corresponding to *Escherichia coli* K12 gene positions, RNA catalytic introns, ORFs and insertions. **c**, Structural models of encoded proteins (1, 2 and 4: coloured by the colours of the rainbow from the amino to the carboxy terminus) and predicted structure for a catalytic RNA intron (3: coloured by base-pairing probability; red is high, green is moderate, and blue is low). Protein Data Bank structures were used as templates for structural modelling (1: accession 1R7M; 2: 1B24; 4: 1B24).

trees (Supplementary Fig. 1), and seen in previous analyses[2–5,11,18]. Phylogenetic analysis defined 35 phyla within the CPR (see later), which encompasses a proposed superphylum, 'Patescibacteria', previously suggested to include just three phyla[5].

The existence of ~1,500 bacterial phyla was recently suggested[19] using a 75% 16S rRNA gene sequence identity threshold. This contrasts with the current view, which includes 29 established phyla and ~60 candidate phyla. Using this recent definition[19], we estimate that the CPR consists of >250 phyla (Fig. 1 and Supplementary Fig. 1). With the addition of >550 Mb of CPR genome sequence, there is sufficient sampling to clearly resolve 14 phyla within the Parcubacteria and 11 phyla within the Microgenomates, which have sufficient sequence divergence to account for >120 and >60 phyla, respectively. We propose that these 25 phyla be recognized because (1) complete and/or draft genomes are available, (2) they are monophyletic lineages in both 16S rRNA gene and concatenated ribosomal protein trees, and (3) they pass an approximate 75% 16S rRNA gene sequence identity threshold. Importantly, regardless of whether

previous phyla designations or new criteria[19] are used, the CPR comprises >15% of domain Bacteria.

A striking finding from analysis of complete and draft genomes (see statistical assessment in Methods) is unusual ribosome composition in CPR bacteria. All CPR and TM6 bacteria lack ribosomal protein L30 (rpL30; Table 1, Extended Data Fig. 8, Supplementary Table 10 and Supplementary Data 4). Apparently non-essential in bacteria[20], this protein is commonly present except in some symbionts, parasites, Cyanobacteria, and throughout the Planctomycetes–Verrucomicrobia–Chlamydiae (PVC) superphylum[21–23]. Although loss of ribosomal protein L25 is often seen in conjunction with absence of rpL30 (ref. 21), TM6 (not within the CPR) is the only candidate phylum studied here for which this is the case. This suggests different trajectories of ribosome evolution between the CPR and other lineages without rpL30.

WS6, WWE3, Saccharibacteria and almost all Microgenomates are missing ribosomal protein L9 (rpL9; Table 1). rpL9 is thought to be universal in bacteria[23], and is involved in both initiation of ribosome

Table 1 | Genomes from candidate phyla bacteria

Lineage	Complete genomes	Draft genomes	Median SCGs	Average genome size in bp (s.d.)	Average per cent GC (min/max)	Missing ribosomal protein(s)
Parcubacteria	2	427	91%	707, 464 (295, 862)	43 (31/60)	L30, OD1-L1, missing L1
Microgenomates	3	252	91%	788, 693 (261, 196)	41 (31/50)	L30, L9*
WWE3	0	41	93%	719, 830 (344, 415)	43 (41/46)	L30, L9
WS6	0	16	91%	584, 741 (167, 526)	34 (33/39)	L30, L9
Peregrinibacteria	0	15	91%	1,183, 124 (344, 415)	42 (33/54)	L30
TM6	0	15	98%	1,060, 264 (167, 526)	36 (28/43)	L30, L25
Berkelbacteria	1	6	88%	581, 936 (243, 398)	39 (34/46)	L30
Kazan	1	5	95%	657, 191 (214, 462)	49 (45/52)	L30
CPR2	0	6	100%	1,032, 375 (183, 809)	39 (38/39)	L30
Saccharibacteria	1	2	99%	971, 756 (157, 794)	47 (46/48)	L30, L9
CPR1	0	2	72%	578, 470 (266, 611)	46 (42/49)	L30
CPR3	0	2	86%	945, 288 (153, 931)	35 (34/35)	L30
All	8	789	91%	749, 453 (263, 507)	42 (28/60)	

The percentage of 43 single copy genes (SCGs) identified in each genome was used to estimate completeness. CPR1–3 are novel CPR lineages.
* One genotype has rpL9.

assembly[24] and maintaining translation fidelity[25], yet culture-based studies suggest it does not contribute to fitness[20]. Of the three complete Microgenomates genomes, one encodes rpL9. This rpL9 sequence is phylogenetically related to Parcubacteria sequences (Supplementary Fig. 1), suggesting acquisition by lateral gene transfer.

Ribosomal protein L1 (rpL1) is absent from a group within the Parcubacteria that potentially includes >90 phyla. We refer to this group as OD1-L1 (Fig. 1). No other organisms are known to lack rpL1, a large protein that forms a prominent feature of the large subunit[26]. This ribosome initiator protein[24] controls its own expression[27], and loss of rpL1 results in severe growth defects[20]. Absence of rpL1 in this diverse clade suggests alternative mechanisms of ribosome regulation, possibly involving an analogous protein and/or an alternative ribosome structure.

The ribosomal protein biogenesis factor GTPase Der is missing from almost all organisms lacking either rpL9 or rpL1 (Extended Data Fig. 8). Der is essential for ribosome production and is conserved throughout bacteria[28]. Thus, in addition to having unusual ribosome composition, many CPR bacteria probably employ alternative ribosome assembly methods. Although some CPR bacteria have both atypical ribosomes and rRNA introns, these features are not directly linked and thus are not compensatory.

Typically, bacteria within a phylum have widely varying genome sizes and metabolic capacities. In contrast, organisms throughout the CPR have consistently small genomes and similar metabolic limitations. Specifically, all have incomplete tricarboxylic acid cycles and lack electron transport chain complexes, including terminal oxidases and reductases; some lack ATP synthase (Extended Data Fig. 8). With the notable exception of the Peregrinibacteria, most have incomplete nucleotide and amino acid biosynthesis pathways. CPR bacteria are probably obligate fermenters dependent on other organisms for survival, although they could support respiring organisms by excreting fermentation end products. Overall, these characteristics, in addition to unusual ribosomes, a high frequency of rRNA introns and a distinct phylogeny, establish the CPR as a subdivision within domain Bacteria.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Harris, J. K., Kelley, S. T. & Pace, N. R. New perspective on uncultured bacterial phylogenetic division OP11. *Appl. Environ. Microbiol.* **70,** 845–849 (2004).
2. Wrighton, K. C. et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337,** 1661–1665 (2012).
3. Kantor, R. S. et al. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *MBio* **4,** e00708–e00713 (2013).
4. Wrighton, K. C. et al. Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. *ISME J.* **8,** 1452–1463 (2014).
5. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499,** 431–437 (2013).
6. Albertsen, M. et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnol.* **31,** 533–538 (2013).
7. Castelle, C. J. et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25,** 690–701 (2015).
8. Luef, B. et al. Diverse, uncultivated ultra-small bacterial cells in groundwater. *Nature Commun.* **6,** 6372 (2015).
9. Burt, A. & Koufopanou, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.* **14,** 609–615 (2004).
10. Salman, V., Amann, R., Shub, D. A. & Schulz-Vogt, H. N. Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. *Proc. Natl Acad. Sci. USA* **109,** 4203–4208 (2012).
11. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–D596 (2013).
12. Evguenieva-Hackenberg, E. Bacterial ribosomal RNA in pieces. *Mol. Microbiol.* **57,** 318–325 (2005).
13. Raghavan, R., Hicks, L. D. & Minnick, M. F. Toxic introns and parasitic intein in *Coxiella burnetii*: legacies of a promiscuous past. *J. Bacteriol.* **190,** 5934–5943 (2008).
14. Baker, B. J., Hugenholtz, P., Dawson, S. C. & Banfield, J. F. Extremely acidophilic protists from acid mine drainage host *Rickettsiales*-lineage endosymbionts that have intervening sequences in their 16S rRNA genes. *Appl. Environ. Microbiol.* **69,** 5512–5518 (2003).
15. Gong, J., Qing, Y., Guo, X. & Warren, A. 'Candidatus Sonnebornia yantaiensis', a member of candidate division OD1, as intracellular bacteria of the ciliated protist *Paramecium bursaria* (Ciliophora, Oligohymenophorea). *Syst. Appl. Microbiol.* **37,** 35–41 (2014).
16. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).
17. Nawrocki, E. P. in *Structural RNA Homology Search and Alignment using Covariance Models* (ed. Eddy, S. R. et al.) (Washington Univ. in Saint Louis, 2009).
18. Baker, B. J. & Dick, G. J. Omic approaches in microbial ecology: charting the unknown. *Microbe* **8,** 353–360 (2013).
19. Yarza, P. et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Rev. Microbiol.* **12,** 635–645 (2014).
20. Akanuma, G. et al. Inactivation of ribosomal protein genes in *Bacillus subtilis* reveals importance of each ribosomal protein for cell proliferation and cell differentiation. *J. Bacteriol.* **194,** 6282–6291 (2012).
21. Lecompte, O. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* **30,** 5382–5390 (2002).
22. Lagkouvardos, I., Jehl, M.-A., Rattei, T. & Horn, M. Signature protein of the PVC superphylum. *Appl. Environ. Microbiol.* **80,** 440–445 (2014).
23. Yutin, N., Puigbò, P., Koonin, E. V. & Wolf, Y. I. Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE* **7,** e36972 (2012).
24. Nowotny, V. & Nierhaus, K. H. Initiator proteins for the assembly of the 50S subunit from *Escherichia coli* ribosomes. *Proc. Natl Acad. Sci. USA* **79,** 7238–7242 (1982).
25. Atkins, J. F. & Björk, G. R. A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol. Mol. Biol. Rev.* **73,** 178–210 (2009).
26. Schuwirth, B. S. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310,** 827–834 (2005).
27. Nevskaya, N. Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Res.* **33,** 478–485 (2005).
28. Shajani, Z., Sykes, M. T. & Williamson, J. R. Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* **80,** 501–526 (2011).

Supplementary Information is available in the online version of the paper.

Author Contributions Samples and geochemical measurements were taken by M.J.W., K.C.W. and K.H.W. B.C.T. assembled the metagenome data. I.S. implemented the ABAWACA algorithm. C.T.B. and J.F.B. binned the data and carried out the ESOM binning validation. J.F.B. closed and curated the complete genomes. C.T.B., L.A.H. and B.C.T. conducted the rRNA gene insertion analysis. C.T.B. and L.A.H. performed phylogenetic analyses. M.J.W. and K.C.W. conducted the RNA sequencing. C.T.B. carried out the 16S rRNA gene copy number, primer binding and transcript analyses. C.T.B. and J.F.B. carried out the ribosomal protein analyses. C.T.B., L.A.H., C.J.C. and J.F.B. conducted the metabolic analysis. A.S. and B.C.T. provided bioinformatics support. C.T.B. and J.F.B. drafted the manuscript. All authors reviewed the results and approved the manuscript.

METHODS

Groundwater sampling and geochemical measurements. We studied ground-water microbial communities from an aquifer adjacent to the Colorado River near Rifle, Colorado, USA at the Rifle Integrated Field Research Challenge (IFRC) site. Aquifer well CD-01 (39° 31′ 44.69″ N, 107° 46′ 19.71″ W; 1,617.5 m above mean sea level) was observed from 23 August to 22 December, 2011, during which a 79-day acetate amendment experiment was conducted (Extended Data Fig. 1 and see refs 7, 8). This well had been subjected to an acetate stimulation experiment during the previous year[29,30]. Acetate (15 μM target concentration within the aquifer) was administered to the alluvial aquifer through a series of injection wells, and micro-bial biomass was sampled from groundwater pumped from a down gradient monitoring well. Approximately 100 l of groundwater was sampled from a depth of 5 m below ground surface through a 1.2 μm pre-filter, and cells were collected on serial 0.2 and 0.1 μm filters (Supor disc filters; Pall Corporation), with the specific objective of enriching for organisms with small cell sizes. Filters were immediately frozen after collection in a dry ice and ethanol bath. See Supplementary Table 1 for sampling dates and the amount of groundwater filtered over the course of the experiment.

Geochemical measurements were made on samples collected 5 m below ground surface (Supplementary Table 1). The Hach phenanthroline assay and sulphide reagent kits were used to measure ferrous iron and sulfide concentrations, respec-tively. Acetate and sulfate concentrations were measured by ion chromatography, as previously described[30]. Briefly, acetate and sulfate concentrations were mea-sured with a Dionex ICS-2100 fitted with an AS-18 guard and analytical column.

Metagenome and metatranscriptome sequencing. Six time points spanning a range of geochemical conditions were chosen for metagenomic and metatran-scriptomic analysis (Extended Data Fig. 1 and Supplementary Table 1). DNA was extracted from ~1.5 g of each frozen filter using the PowerSoil DNA Isolation Kit (MO-BIO Labs) with the following modifications: DNA was con-centrated by sodium acetate/ethanol precipitation with glycogen, and DNA was eluted in 50 μl Tris buffer. DNA library preparation and sequencing was con-ducted at the Joint Genome Institute. Total DNA was sequenced on an Illumina HiSeq, producing 150 bp paired reads with a targeted insert size of 500 bp. Sequence data were processed using version 1.8 of the Illumina CASAVA pipeline, and all reads were trimmed based on quality scores using Sickle (https://github.-com/najoshi/sickle; default parameters; Supplementary Table 2).

RNA was extracted from the 0.2 μm filters using the Invitrogen TRIzol reagent, followed by genomic DNA removal and cleaning using the Qiagen RNase-Free DNase Set kit and the Qiagen Mini RNeasy kit. An Agilent 2100 Bioanalyzer (Agilent Technologies) was used to assess the integrity of the RNA samples. The Applied Biosystems SOLiD Total RNA-Seq kit was used to generate the cDNA template library. The SOLiD EZ Bead system (Life Technologies) was used to perform emulsion clonal bead amplification to generate bead templates for SOLiD platform sequencing. Samples were sequenced at Pacific Northwest National Laboratory on the 5500XL SOLiD platform. The 50 bp single reads were trimmed using Sickle (default parameters; Supplementary Table 2).

Metagenome assembly, annotation and genome binning. Total community DNA was assembled individually for each sample using IDBA_UD[31] with default parameters (Supplementary Table 2). 16S and 23S rRNA gene sequences were identified from all assembled sequences and curated using an automated method (see later). Scaffold coverage was calculated by mapping reads back to the assembly using Bowtie2 (ref. 32) with default parameters for paired reads. All scaffolds ≥5 kb were included when binning genomes from the metagenome assembly. These scaffolds were annotated by first predicting ORFs using the metagenome implementation of Prodigal[33], and then using USEARCH (-ublast; http://drive5.com/usearch/)[34] to search protein sequences against UniRef90 (ref. 35), KEGG[36,37], and an in-house database composed of ORFs predicted from genomes of candid-ate phyla organisms. The in-house database includes previously published gen-omes[2–4,38,39] and genomes from ongoing work. Scaffolds were binned on the basis of their GC content, DNA sequence coverage, abundance pattern across samples and taxonomic affiliation, both automatically with the ABAWACA algorithm (see later) and manually using ggKbase (http://ggkbase.berkeley.edu/). Bins generated by ABAWACA were manually inspected within ggKbase. Reported here are gen-omes binned for organisms associated with the CPR (Fig. 1 and Supplementary Table 3) and TM6 (a phylum of organisms with similar characteristics).

To test the accuracy of this binning method, 20 draft-quality genomes were randomly selected from a sample with a high proportion of CPR genomes (GWA2). These genomes were fragmented and then re-clustered on the basis of tetranucleotide signatures using an emergent self-organizing map (ESOM), as previously described[40]. Tetranucleotide frequencies were calculated for 5–10 kb scaffold fragments. The number of occurrences of each tetranucleotide in each fragment was normalized on the basis of the total number of times the tetranucleotide was observed across all fragments, and then these values were

log-transformed, standardized so they would follow a normal distribution, and then scaled from 0–1. Normalized tetranucleotide values for each fragment were standardized so that they would also follow a normal distribution. The resulting matrix was used to train an ESOM for 100 epochs using esom_train.pl (https://github.com/micronorman/bantools) (downloaded October 2014). The ESOM was visualized using the Databionic ESOM Tools software[62] (http://databionic-esom.sourceforge.net/). Colouring fragments (data points) in the ESOM on the basis of the genome each fragment originated from enabled validation of these genome bins (Extended Data Fig. 2).

ABAWACA genome binning. ABAWACA was used to generate preliminary genome bins for each sample. This algorithm assesses different characteristics of assembled scaffolds to bin them into genomes. Here, we used a combination of mono-, di- and tri- nucleotide frequencies and coverage values calculated by mapping DNA sequences from all samples to the scaffolds from the sample being binned. This algorithm uses the given information in a hierarchical clustering fashion as follows. First, all scaffolds are broken into 5 kb segments called data points, and the properties of each data point are computed. The binning process begins with a single bin that contains all scaffolds and proceeds by iteratively splitting this and subsequent bins. All non-final bins are evaluated during each iteration. The algorithm searches for a single value for one of the characteristics that will result in the best separation of the scaffolds into two bins. Separation quality is calculated based on the number of data points that were assigned cor-rectly given the separation of the scaffolds. Once a split has been made, scaffolds are separated into the bin with the majority of the data points representing the scaffold. Bins are approved if the quality score exceeds a predefined threshold, and both bins consist of at least 50 data points. A bin is considered final if no separation can be made; otherwise, it undergoes further rounds of binning.

Genome assessment and finishing. Genome bins were associated with CPR lineages on the basis of phylogenetic analysis of 16S rRNA genes and/or ribosomal proteins (see later). When these phylogenetic markers were not present for a particular genome bin, taxonomic placement was achieved based on a consensus of the taxonomic assignments given to ORFs on the basis of their similarity to ORFs from CPR representatives in the candidate phyla database described earlier. Genome completeness was assessed using a modified version of a previously reported list of universal single copy genes (SCGs) for bacteria[41] (Supple-mentary Table 3). Several SCGs were not included as they were found to be unsuitable for the CPR, either because these genes were too divergent in CPR genomes to be reliably detected, or because members of the CPR do not encode these genes. For example, the genes for ribosomal proteins L1 and L9 are not encoded in the genomes of many CPR organisms (see main text). SCGs were identified based on a reciprocal best BLAST[42] hit procedure using a database of SCG protein sequences from a representative set of genomes. First, SCG proteins from the database were searched against all protein sequences in a given genome to identity SCG candidates (blastall -p blastp -F F -e 1e-2). Then, these candidate proteins were searched against the SCG protein sequence database to confirm the assignment (blastall -p blastp -F F -e 1e-5 -b 1 -v 1). SCGs were considered to be present if they were identified by the reciprocal hit method, and the best alignment with a database sequence covered ≥50% of the protein sequence.

To be included in this study as a draft genome, a bin must have contained at least 50% of these SCGs with fewer than 1.125 copies of the genes (indicating that the bin does not contain appreciable contamination from other genomes). To make consistent comparisons with previously sequenced genomes from the CPR, all avail-able genomes were re-assessed using these methods (Supplementary Table 4)[43–45].

Several high-quality genome bins were selected for manual curation and genome finishing. Binned scaffolds were connected with one another by extending scaffolds and searching for overlaps. Scaffold extension was achieved by assembling reads mapped to the ends of scaffolds. Assembly errors were detected by manually inspect-ing the read mapping for these genomes. Genomes were only considered to be complete if they were circular, did not contain gaps, and were, based on complete visual inspection of mapped reads, free of assembly errors. Assembly errors can be identified as regions that do not have read support (that is, reads may map but with mismatches, or regions may not be supported by paired reads). These regions can be manually corrected. Genomes were also checked for the presence of 'orphaned pairs', which could indicate alternative assembly paths. The complete genome for GWB1_sub10_OD1-complete was obtained by first assembling 1/10 of the sequence data for sample GWB1, binning scaffolds on the basis of GC content, coverage and taxonomic affiliation, and then genome finishing as described earlier.

Identification of rRNA genes and insertions. 16S and 23S rRNA gene sequences were identified based on hidden Markov model (HMM) searches using the cmsearch program from the Infernal package[46] (cmsearch -hmmonly -acc -noali -T -1). Importantly, all identified gene sequences were curated to remove assembly errors before any analysis was conducted (see later). To identify 16S rRNA gene sequences, all assembled contigs were searched against the manually

curated structural alignment of the 16S rRNA provided with SSU-Align[17]. Since the SSU-Align 16S rRNA gene covariance model did not include sequences with insertions, large gaps in the alignment between each sequence and the model revealed the boundaries of insertions. Because no equivalent model existed for the 23S rRNA gene, we built a sequence-only model from the manually curated seed alignment maintained by the Comparative RNA Web[47] (Supplementary Data 3). While this model did not contain secondary structure information, it was appropriate for identifying 23S rRNA genes, and the boundaries of insertion sequences, from sequence-based HMM alignments, as was done for 16S rRNA genes. To identify the location of rRNA gene insertions with respect to well-studied *Escherichia coli* sequences, all bacterial rRNA gene sequences found to encode insertions were aligned against models consisting of only the respective rRNA from *E. coli* strain K12 substrain DH10B (Fig. 2, Extended Data Fig. 6 and Supplementary Table 5).

Similarity of rRNA insertions to previously studied structural RNA families (for example, group I and group II catalytic RNAs) was determined by searching full rRNA sequences against Rfam[48] using cmscan (also from Infernal; Supplementary Table 5). Regions of the rRNA with significant alignments to a structural RNA family (passed model inclusion threshold) were considered as positive hits if at least 25% of the alignment overlapped with an insertion. These rRNA structural families were of particular interest for determining whether or not insertions encode catalytic RNAs potentially capable of self-splicing from containing RNA sequences (Fig. 2 and Extended Data Fig. 6). RNA secondary structure was predicted for selected intervening sequences using the Andronescu 2007 model[49] implemented in Geneious v. 7.1.5 (ref. 50) (Fig. 3).

ORFs encoded within rRNA insertion sequences were identified by first predicting ORFs across full rRNA genes, and then selecting ORFs encoded within insertion regions. To exclude false ORF predictions, at least 90% of the ORF had to overlap with an insertion. Insertion-encoded ORFs were searched against Pfam[51] to associate encoded proteins with known families (Fig. 2, Extended Data Fig. 6 and Supplementary Table 5). In some cases, Phyre2 (ref. 52) was used to model protein sequences and provide further support for identified homing endonucleases (Fig. 3). Insertions and ORFs identified within 16S and 23S rRNA genes were compared with one another using BLAST (Supplementary Table 9). To assess the prevalence and types of intervening sequences previously sampled in 16S rRNA genes from bacteria, version 115 of non-redundant SILVA[11] was analysed using the same methods (Extended Data Fig. 4 and Supplementary Table 6). Importantly, all insertions ≥10 bp were removed before multiple sequence alignment and phylogenetic analysis of 16S rRNA gene sequences.

Bacterial community composition based on assembled 16S rRNA genes. The composition of the bacterial community was determined on the basis of assembled and curated 16S rRNA gene sequences. Each sequence was given a taxonomic assignment based on the phylogenetic analysis described later. Coverage of all assembled 16S rRNA gene sequences was determined for each sample by stringently mapping reads using Bowtie2 (no mismatches allowed). For each sample, the coverage of all sequences belonging to each lineage of interest was summed, and then converted to a percent relative abundance to observe the composition of each filtrate and shifts in the community across the time series (Extended Data Fig. 3).

16S rRNA gene copy number. 16S rRNA gene copy number was estimated for all complete and draft genomes based on two assessments. First, the number of assembled 16S rRNA gene sequences was determined. Second, coverage of 16S rRNA gene regions was compared with the coverage of the rest of the genome to determine relative copy number. Relative copy number was calculated because of the likeliness of assembling only one 16S rRNA gene for organisms with multiple, identical copies of the gene. Owing to the conserved nature of the 16S rRNA gene, it is common for these regions to have inflated coverage values based on default mapping parameters due to inaccurate assignment of reads to sequences from other organisms. To avoid this, both genome and 16S rRNA gene coverage values were calculated based on reads that mapped with zero mismatches. Relative copy number was calculated as: (16S rRNA gene coverage)/(genome coverage). Copy number for each genome was determined by whichever value was greatest, the number of assembled genes or relative copy number (Extended Data Fig. 5 and Supplementary Table 7). Only ten CPR genomes were found to encode more than one copy of the 16S rRNA gene; however, since these genes were not similar to one another, it is more likely that these rare cases were binning errors.

rRNA gene transcript analysis. To determine the fate of rRNA insertion sequences, RNA transcript sequences recovered from 0.2 μm filters were stringently mapped to assembled, curated rRNA genes. To prevent short reads from erroneously matching to either rRNA genes or insertions, zero mismatches were allowed between reads and assemblies. Coverage was calculated separately for 16S rRNA gene and predicted insertion regions, and then the values were compared with one another (Supplementary Table 8). Most insertions were found to have

zero coverage. However, in some cases very low coverage of insertion regions was found. In almost all cases these low coverage values were the result of a small portion of the insertion region being covered by RNA sequence, probably the result of a small difference between predicted and actual insertion regions, but possibly the result of partial recovery of spliced insertion sequences.

16S rRNA gene primer binding analysis. The level of sequence divergence of the 16S rRNA genes assembled here from metagenome data, compared with sequences from existing databases, suggests that they would elude PCR-based analysis. We assessed the binding affinity of commonly used 16S rRNA gene survey primers 515F and 806R[16,53]. Assembled 16S rRNA gene sequences were clustered at 97% sequence identity using USEARCH (-cluster_smallmem -query_cov 0.50 -target_cov 0.50 -id 0.97) to remove redundant sequences from the analysis. Because some of the sequences are not complete, only those spanning the 515–806 region of the *E. coli* 16S rRNA gene were included. Primer binding was assessed with PrimerProspector[54] using default parameters (Extended Data Fig. 7).

Phylogenetic analysis. Phylogenetic analysis was carried out using several different marker sequences in order to best survey the diversity within the groundwater microbial community, and to robustly assign taxonomy to complete and draft genomes. Markers included the 16S rRNA gene, ribosomal proteins encoded by a syntenic block of genes, and ribosomal protein S3 (rpS3). The syntenic block encodes the genes for ribosomal proteins L2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24 and S3, 8, 10, 17, 19, hereafter referred to as rp16. In the rp16 analysis, individual protein sequence alignments were concatenated for phylogenetic inference. Unlike in previous metagenomic studies, near-complete 16S rRNA gene sequences were assembled commonly enough to be able to infer phylogeny for many community members. However, rp16 was also used for phylogenetic analysis because (1) it is encoded in genomes as a syntenic block and is found in only one copy, and thus can be used as a proxy for a particular genotype independent of binning, (2) it encodes ribosomal proteins that provide a robust phylogenetic signal, and (3) it is assembled more frequently from metagenome sequence data compared with the 16S rRNA gene[38]. rpS3 was also independently used as a phylogenetic marker because of its strong phylogenetic signal, despite having a relatively short protein sequence. In cases where a genome did not contain any of these markers (Supplementary Table 3), taxonomic assignment was made based on whole genome comparisons to the database of reference genomes described earlier. In all cases, metagenome assembly was necessary for providing a robust phylogenetic analysis.

After removing insertions ≥10 bp from 16S rRNA gene sequences from this and previous studies, sequences were aligned with SSU-Align. SSU-Align classifies sequences as bacteria, archaea or eukarya, and then generates separate alignments for sequences from each domain. The resulting Stockholm-formatted bacterial multiple sequence alignment was converted to FASTA, and all alignment insert columns were removed. This resulted in a 1,582 bp alignment. All sequences with ≥800 bp of aligned sequence were used for phylogenetic analysis. Several archaeal reference sequences were chosen for the phylogenetic root, aligned to the bacterial 16S rRNA gene model provided with SSU-Align, and concatenated with the bacterial multiple sequence alignment. A maximum-likelihood phylogeny was inferred using RAxML[55] with the GTRCAT model of evolution and 100 bootstrap re-samplings (Supplementary Fig. 1 and Supplementary Data 2). A subset of the tree was annotated using GraPhlAn (http://huttenhower.sph.harvard.edu/graphlan) (Fig. 1).

rp16 ORFs were identified by searching all ORFs encoded on scaffolds ≥5 kb against databases of each of these ribosomal proteins. Searches were carried out with USEARCH (-ublast). Syntenic groups of ORFs were selected if at least three of the ribosomal proteins in rp16 could be identified with an E-value ≤1 × 10⁻⁶. This allowed for identification of all instances of each ribosomal protein in rp16 encoded within assembled scaffolds. For each ribosomal protein, all identified protein sequences along with reference sequences were aligned to their respective Pfam HMM profile using hmmalign from the HMMER 3.0 package[56]. Protein sequence alignments were converted from Stockholm format to FASTA, alignment insert columns were removed, and the 16 protein alignments concatenated. This resulted in a 1,935 amino acid alignment. All sequences with ≥1,000 aligned residues were kept for phylogenetic analysis. Because of the size of the multiple sequence alignment, phylogenetic analysis was carried out in two steps. First, FastTree2 (ref. 57) was used to infer the phylogeny of the entire sequence set using the Jones–Taylor–Thornton model of amino acid evolution (JTT) and by assuming a single rate of evolution for each site, the 'CAT' approximation (additional options: -spr 4 -mlacc 2 -slownni). Then, sequences associated with the CPR and TM6 were selected, along with representatives of the Archaea and Chloroflexi, to infer a maximum-likelihood phylogeny using RAxML with the LG + alpha + gamma model of evolution and 100 bootstrap re-samplings (see ref. 38 for choice of evolutionary model). Archaea were included as a root for the tree, and Chloroflexi as a root for the CPR. Notably, the CPR is evident as a monophyletic

group in both of these analyses, and in the 16S rRNA gene phylogeny (Fig. 1 and Supplementary Fig. 1).

Phylogenies were inferred from individual protein sequences for rpS3 and ribosomal protein L9 (rpL9). All rpS3 protein sequences were identified from metagenome ORFs by searching protein annotation descriptions. The same was done for rpL9, except only sequences associated with CPR genome bins were included. Erroneously annotated sequences were excluded based on the alignment score inclusion threshold for their respective Pfam HMM profiles (aligned using hmmalign), followed by manual removal of non-rpS3 or rpL9 sequences. Sequences were combined with reference sequences and aligned. rpS3 sequences were aligned to Pfam HMM profile PF00189 using the same procedure as was described for the rp16 protein sequences (see earlier). rpL9 was aligned using MUSCLE[58]. All sequences with ≥50 aligned amino acid residues were used for phylogenetic analysis using RAxML with 100 bootstrap re-samplings and an evolutionary model chosen using ProtTest[59] (Supplementary Fig. 1). The ProtTest 2.4 server[59] was run on the Pfam seed alignment for rpS3 and on a random subset of the rpL9 alignment, indicating that the LG + gamma, and the LG + gamma (with fixed base frequencies) evolutionary models should be used for rpS3 and rpL9, respectively.

All phylogenetic trees were visualized using Dendroscope[60].

Identification of novel phyla. The number of phyla within the CPR, Parcubacteria (OD1), and Microgenomates (OP11) was estimated by counting 16S rRNA gene sequence clusters created based on a 75% sequence identity threshold. After removing insertions ≥10 bp, sequences were clustered using USEARCH (-cluster_smallmem -query_cov 0.50 -target_cov 0.50 -id 0.75). This threshold and method for estimating the number of phyla were proposed previously[19]. These authors proposed that phyla could be identified as monophyletic lineages composed of members distinguished by approximately this level of sequence divergence. We classified new phyla based on this and additional, strict criteria. Clusters of 16S rRNA genes that share ≥75% sequence identity were used to assess the divergence and coherence of deep branches of the phylogenetic tree (Supplementary Fig. 1). Bootstrap support values were often higher for lineages primarily composed of one or few clusters, validating the use of this threshold. Lineages were proposed as phyla if (1) they formed a monophyletic group in the 16S rRNA gene phylogeny, (2) 16S rRNA genes were approximately 25% divergent from other lineages, (3) they were also supported by the rp16 concatenated ribosomal protein phylogeny, and (4) representative complete and/or draft genomes were available. Names for these phyla were proposed based on the names of lifetime achievement award recipients in microbiology (Fig. 1, Extended Data Table 1 and Supplementary Fig. 1). Genomes were associated with these phyla using the 16S rRNA gene and/or rp16 phylogenies (Supplementary Table 3).

Sequence curation. Assembled 16S rRNA genes, 23S rRNA genes, and scaffold regions encoding rp16 genes were curated to identify and fix assembly errors before assessment of insertions in rRNA genes and/or phylogenetic analysis. For curation, these genes were extracted along with 2 kb of sequence from each side. Assembly errors, typically short regions of misassembled sequence associated with scaffolding contigs with one another, were identified as regions with zero coverage by stringently mapped paired-end reads. Only one mismatch per read was permitted and only paired reads were included in the analysis. Regions with 1× coverage were only allowed if at least 3 bp on either side of the read overlapped with other reads, with zero mismatches in the overlap region. When an assembly error was detected, read pairs mapped (Bowtie2) to a 1 kb region surrounding the error were collected and reassembled using Velvet[61]. Reads were collected for reassembly as long as at least one read in the pair mapped with two or fewer mismatches. Velvet was run by iterating from kmer 21 to 71, increasing by 10 in each iteration. Reassembled fragments were then merged with the original assembly based on overlap of ≥10 bp. All assembly modifications were verified with a subsequent round of error detection. If an error could not be corrected, the original scaffold was split at the position of the error. In addition to error correction, reads mapped to the ends of scaffolds were reassembled and used to extend scaffolds, or the ends of broken scaffolds, when possible. After curation, genes of interest were re-identified on curated scaffolds using the methods described earlier (Supplementary Data 1). On average, 1.5 assembly errors were corrected for each scaffold region containing a 16S rRNA gene.

Ribosomal protein inventory and metabolic potential of CPR genomes. Metabolic potential of CPR genomes was assessed using ggKbase. In ggKbase, lists related to different proteins or metabolic pathways were generated by searching for specific keywords in gene annotations. Here, lists were created to assess ribosomal protein composition and metabolic potential across the CPR (Extended Data Fig. 8). Genomes were compared with one another by creating ggKbase genome summaries based on a selection of these lists. This allowed for the simultaneous assessment and comparison of the 8 complete and 789 draft-quality genomes assembled here.

To compare genomes on the basis of both their phylogenetic associations and metabolic capacity, and to get the clearest picture of the metabolic potential of the CPR, an additional analysis was conducted with only complete and near-complete genomes (≥75% of single copy genes and ≤1.125 copies, including an assembled 16S rRNA gene). Since similar genotypes were assembled independently from different samples, this set of complete and near-complete genomes was de-replicated by choosing a representative genome for all flat branches on the 16S rRNA gene tree (Supplementary Fig. 1). The genome summary was then ordered based on the 16S rRNA gene phylogeny, a step that was critical for identifying lineages missing specific ribosomal proteins (Extended Data Fig. 8). To find ribosomal proteins that may have evaded detection due to sequence divergence, six-frame translations (bacterial translation table 11) of all complete and draft CPR genomes were searched against Pfam ribosomal protein HMM profiles using hmmscan; however, this confirmed the initial finding of missing ribosomal proteins in organisms from CPR lineages (Supplementary Table 10).

Although complete genomes are invaluable for metabolic analyses, this extensive inventory of draft-quality genomes from organisms representing diverse lineages, and assembled from different samples, enabled confident assessment of gene absence. For example, there are no reported complete WS6 genomes, but the 16 reconstructed draft-quality genomes from this study (median estimated completeness of 91%) showed that this lineage is missing rpL9. The probability of the gene being present, but missing in all 16 genome reconstructions, is $(1 - 0.91)^{16}$, that is, $\sim 2 \times 10^{-17}$. Even if we lower the completion requirement to a very conservative value of 35% complete, 16 such genomes would yield a confidence value of 0.001 for the gene being absent. For lineages where we have hundreds of genomes the probability of missing the gene due to chance is effectively zero.

Code availability. ABAWACA is maintained under https://github.com/CK7/abawaca (version 1.00 used in this analysis: https://github.com/CK7/abawaca/releases/tag/v1.00) and the script used for curating scaffolds, re_assemble_errors.py, is maintained under https://github.com/christophertbrown/fix_assembly_errors (version 1.00 used in this analysis: https://github.com/christophertbrown/fix_assembly_errors/releases/tag/1.00).

29. Luef, B. et al. Iron-reducing bacteria accumulate ferric oxyhydroxide nanoparticle aggregates that may support planktonic growth. ISME J. **7,** 338–350 (2013).
30. Williams, K. H. et al. Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater. Geomicrobiol. J. **28,** 519–539 (2011).
31. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28,** 1420–1428 (2012).
32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nature Methods **9,** 357–359 (2012).
33. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11,** 119 (2010).
34. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26,** 2460–2461 (2010).
35. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics **23,** 1282–1288 (2007).
36. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. **40,** D109–D114 (2012).
37. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **28,** 27 (2000).
38. Hug, L. A. et al. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome **1,** 22 (2013).
39. Castelle, C. J. et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nature Commun. **4,** 2120 (2013).
40. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. Genome Biol. **10,** R85 (2009).
41. Raes, J., Korbel, J. O., Lercher, M. J., von Mering, C. & Bork, P. Prediction of effective genome size in metagenomic samples. Genome Biol. **8,** R10 (2007).
42. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. Basic local alignment search tool. J. Mol. Biol. **215,** 403–410 (1990).
43. McLean, J. S. et al. Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. Proc. Natl Acad. Sci. USA **110,** E2390–E2399 (2013).
44. Podar, M. et al. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl. Environ. Microbiol. **73,** 3205–3214 (2007).
45. Marcy, Y. et al. Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc. Natl Acad. Sci. USA **104,** 11889–11894 (2007).
46. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. Bioinformatics **25,** 1335–1337 (2009).
47. Cannone, J. J. et al. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics **3,** 2 (2002).

48. Burge, S. W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41,** D226–D232 (2013).

49. Andronescu, M., Condon, A., Hoos, H. H., Mathews, D. H. & Murphy, K. P. Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics* **23,** i19–i28 (2007).

50. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

51. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).

52. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4,** 363–371 (2009).

53. Gilbert, J. A. *et al.* Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Stand. Genomic Sci.* **3,** 243–248 (2010).

54. Walters, W. A. *et al.* PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* **27,** 1159–1161 (2011).

55. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

56. Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput. Biol.* **7,** e1002195 (2011).

57. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490 (2010).

58. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

59. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21,** 2104–2105 (2005).

60. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61,** 1061–1067 (2012).

61. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).

62. Ultsch, A. & Moerchen, F. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. Technical Report no. 46 (Dept. of Mathematics and Computer Science, University of Marburg, Germany, 2005).

a
Collection Site: Rifle IFRC

● GW2011 - well CD-01 - 2011

Colorado River

scale in meters
200 100 0 200
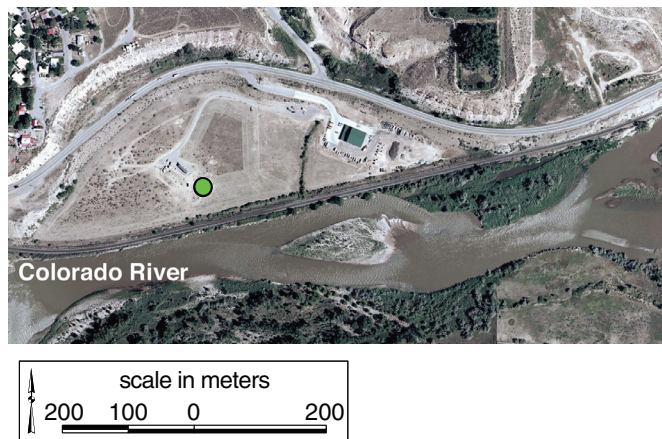
Sample Collection and Filtering

acetate-stimulated groundwater (CD-01 - 2011)

acetate

pump

vadose zone

~3 m

aquifer

groundwater flow

injection well   sampling well

aquitard

1.2 μm
0.2 μm
0.1 μm
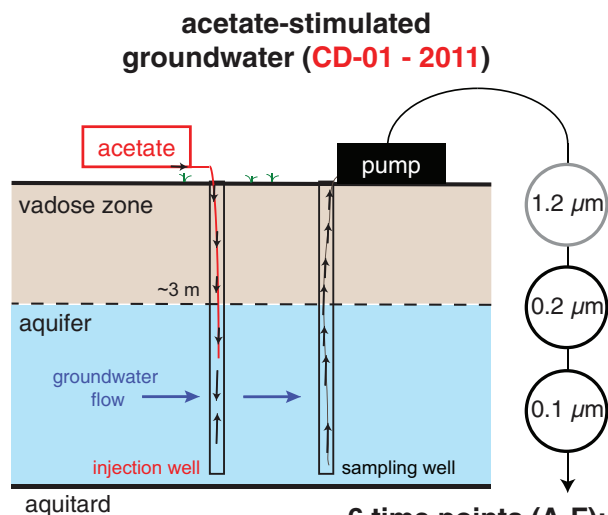
6 time points (A-F):
DNA and RNA extracted
from 0.1 and 0.2 μm filters

b
Geochemical Measurements and Biological Sampling

days following acetate ammendment
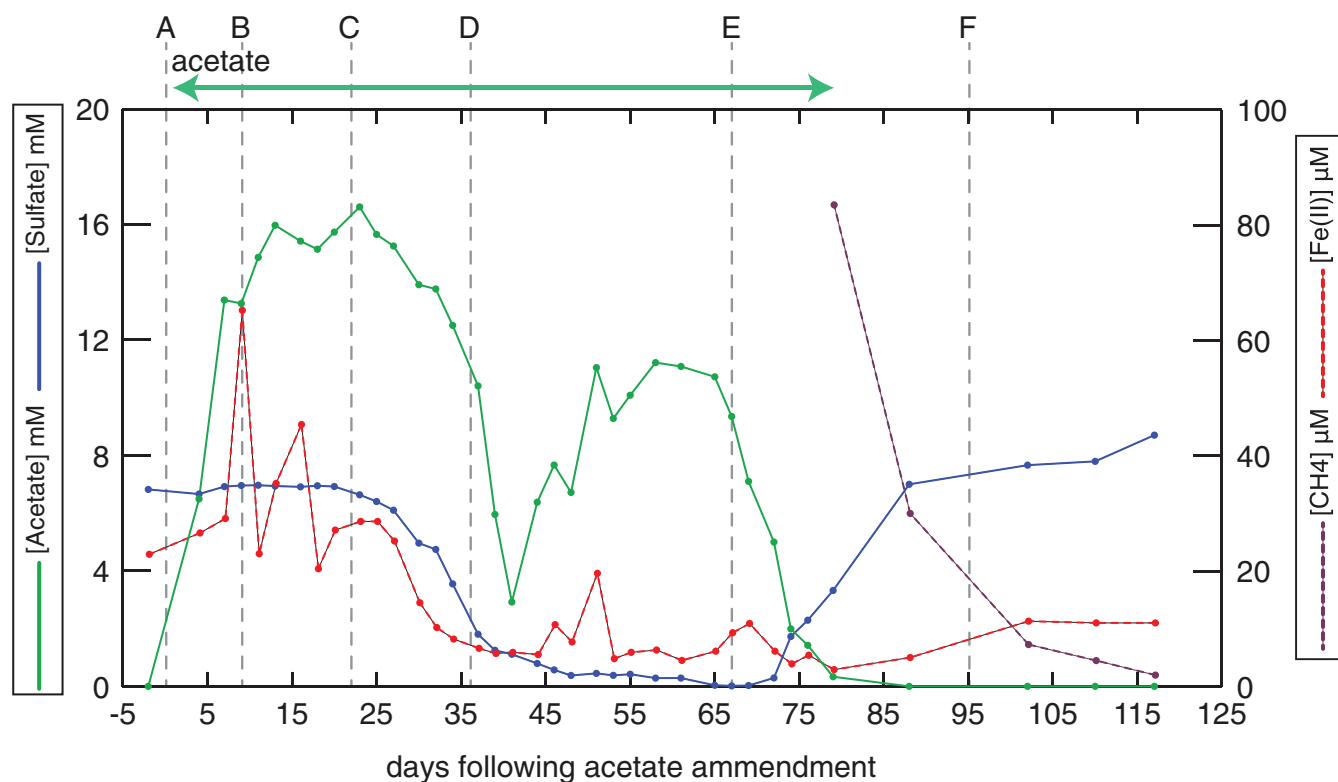
Extended Data Figure 1 | Sampling and geochemical measurements from acetate amendment field experiment conducted in aquifer well CD-01 at the Rifle IFRC site. a, b, Samples were collected for metagenomics and metatranscriptomics at six time points (A–F) spanning several redox transitions during acetate stimulation of groundwater microbial communities. **a,** Groundwater was pumped from the alluvial aquifer and filtered through serial 1.2, 0.2 and 0.1 μm filters. DNA was extracted and sequenced from both the 0.2 and 0.1 μm filters, and RNA extracted and sequenced from the 0.2 μm filters (aerial image provided by S. M. Stoller for the US DOE under contract DE-AM01-07LM00060). **b,** Geochemical measurements were taken throughout the time series, showing a transition from dominant iron reduction to sulfate reduction through to methane production in the sampling environment.

1 GWA2_ACD58_46_7		11 GWA2_OD1_53_7_partial	
2 GWA2_OD1_33_14		12 GWA2_OD1_rel_42_14	
3 GWA2_OD1_41_55_partial		13 GWA2_OP11_33_20	
4 GWA2_OD1_42_41_partial		14 GWA2_OP11_34_18_partial	
5 GWA2_OD1_43_13		15 GWA2_OP11_40_7b	
6 GWA2_OD1_43_66		16 GWA2_OP11_43_14	
7 GWA2_OD1_46_10		17 GWA2_OP11_44_7	
8 GWA2_OD1_46_7_partial		18 GWA2_OP11_47_11b	
9 GWA2_OD1_49_16_partial		19 GWA2_OP11_47_70_partial	
10 GWA2_OD1_50_10_part		20 GWA2_PER_33_10	

Extended Data Figure 2 | Validation of 20 draft-quality genomes by ESOM clustering of genome fragments based on tetranucleotide sequence composition. For validation, 20 draft genomes from a sample with a high proportion of CPR genomes (GWA2) were chosen at random. Each data point represents a 5–10 kb genome fragment. The ESOM was trained for 100 epochs with normalized tetranucleotide frequencies. Dark lines between data points indicate strong separation between regions. Data points are coloured based on the genome the fragment originated from. The ESOM shows well-delineated clusters for most of the 20 draft genomes, with few sequence fragments falling outside of these clusters. Two genomes from the same Microgenomates (OP11) phylum were not well delineated in the tetranucleotide-based ESOM (genomes 18 and 19). This shows how the method we used for binning, which takes into account abundance patterns in addition to sequence signatures, provides more accurate genome reconstructions. The white box distinguishes a single period on the repeating map. Genomes split into multiple clusters are labelled in red.

a

b

Legend:
- Other Bacteria
- Microgenomates (OP11)
- Berkelbacteria
- Saccharibacteria (TM7)
- SM2F11
- Parcubacteria (OD1)
- TM6
- Peregrinibacteria (PER)
- Gracilibacteria (BD1-5)
- Kazan
- WWE3
- WS6
- CPR2
- CPR3
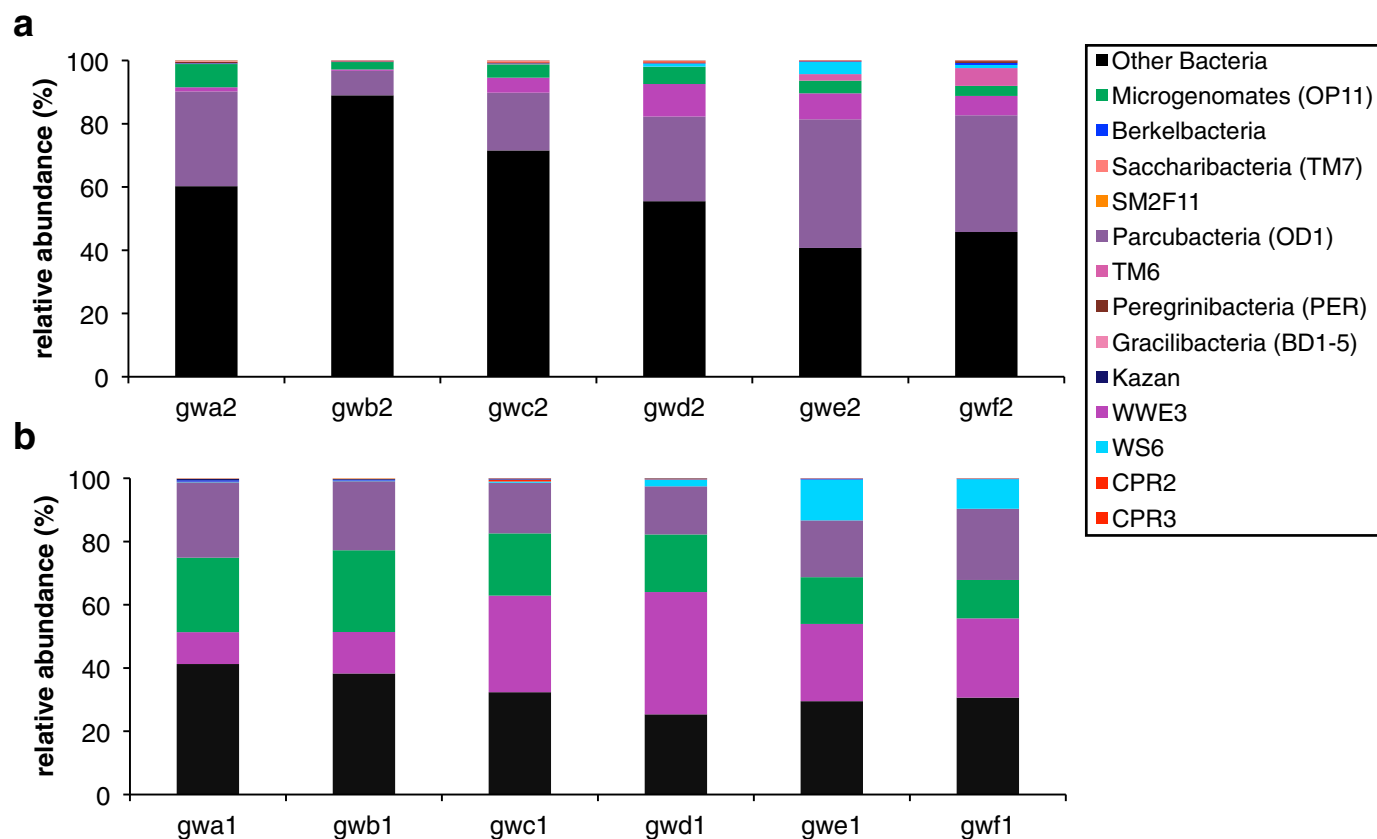
Extended Data Figure 3 | Relative abundance of bacterial community members during acetate amendment. a, b, Relative abundance was calculated based on stringent mapping of paired-read sequences from each sample to 16S rRNA gene sequences assembled from all samples. Relative abundance of cells from 0.2 μm filters (**a**) and from 0.1 μm (**b**) filters. Enrichment of CPR organisms in the 0.2 μm filtrate indicates that these organisms have ultra-small cell sizes.
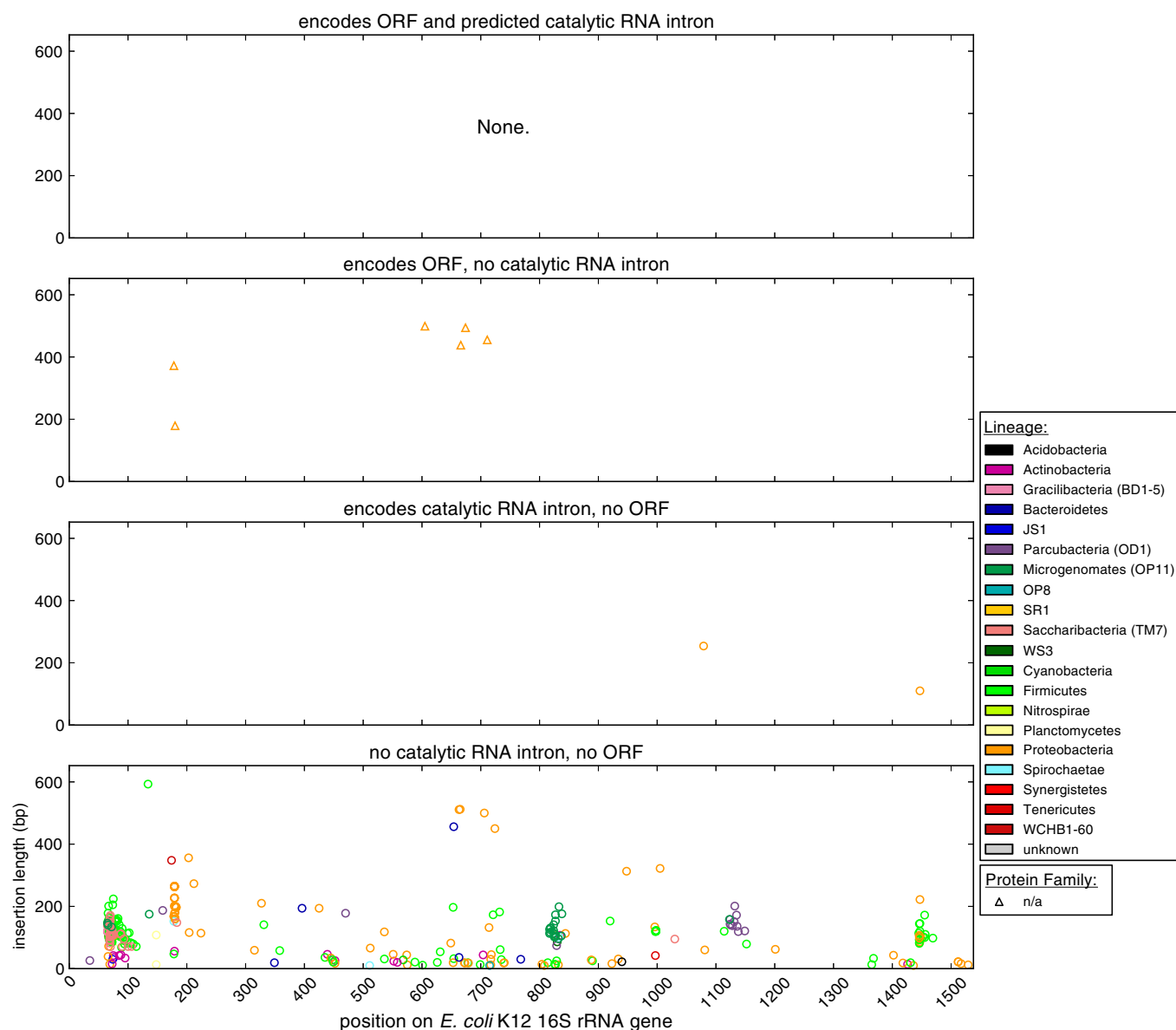
Extended Data Figure 4 | Features of insertion sequences encoded within 16S rRNA genes from the Silva database. The non-redundant Silva 16S rRNA gene database (v. 115) was analysed to assess the prevalence of insertions. Only 761 of the 418,498 16S rRNA gene sequences from bacteria encode insertions. While many small insertions were identified, unlike the 16S rRNA gene sequences assembled from groundwater, these sequences (1) rarely encode large insertions, (2) do not contain both ORFs and introns, (3) do not encode ORFs that could be assigned to Pfam families, and (4) may be found in one of multiple copies of the 16S rRNA gene.
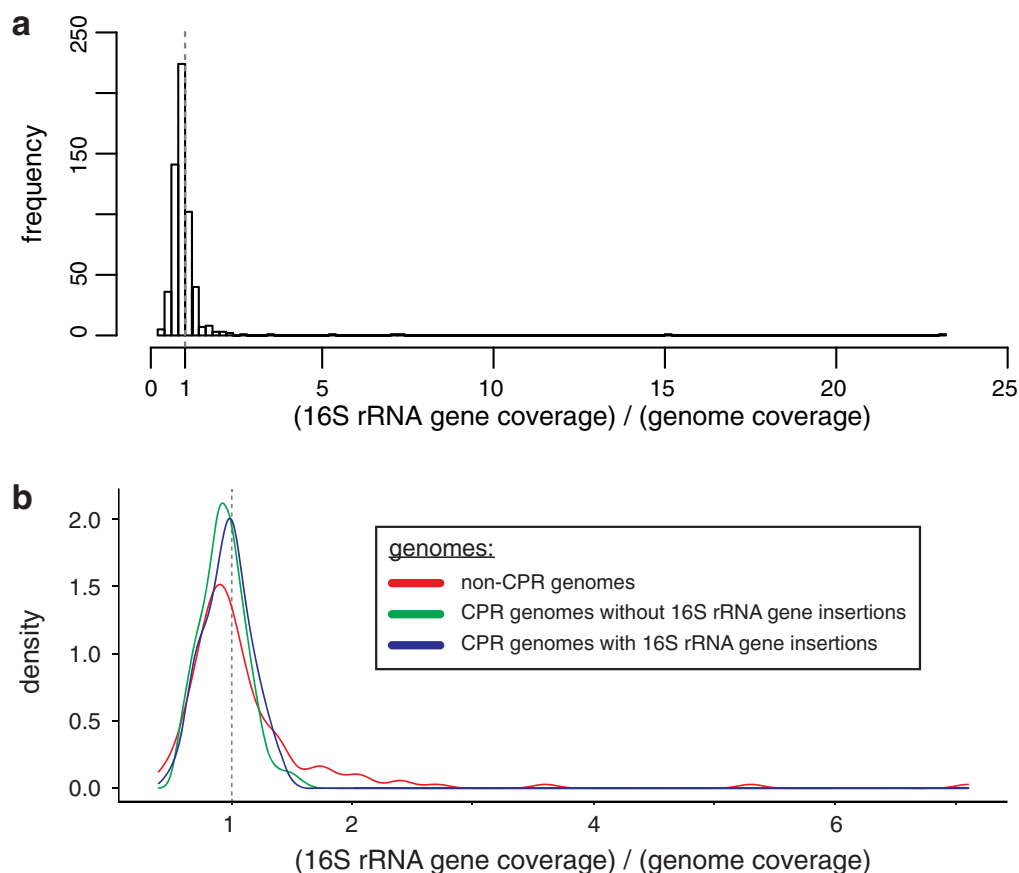
Extended Data Figure 5 | 16S rRNA gene copy number estimations for genomes reconstructed from groundwater metagenomics. a, b, 16S rRNA gene copy number was estimated for all draft CPR genomes and genome bins for organisms outside the CPR. This was achieved by comparing the coverage of 16S rRNA gene regions to the coverage of the rest of the genome. Importantly, coverage was calculated only with stringently mapped reads (no mismatches were allowed) to improve the accuracy of coverage calculations. **a,** Histogram of the number of 16S rRNA gene sequence copies estimated for each genome by calculating (16S rRNA gene coverage)/(genome coverage). Several WWE3 genomes were estimated to have high 16S rRNA gene copy number (Supplementary Table 7), but it was later determined that these estimates were skewed by the presence of a highly abundant closely related strain. The complete WWE3 genome assembled previously[3] has an identical 16S rRNA gene and confirms that it is found in only one copy for this genotype. Thus, we removed these estimates from subsequent copy number analysis. **b,** Density plot comparing estimated copy number of genomes for organisms found within and outside the CPR, where the longer tail for non-CPR genomes depicts the propensity for multiple 16S rRNA copies, a trait absent from the CPR.
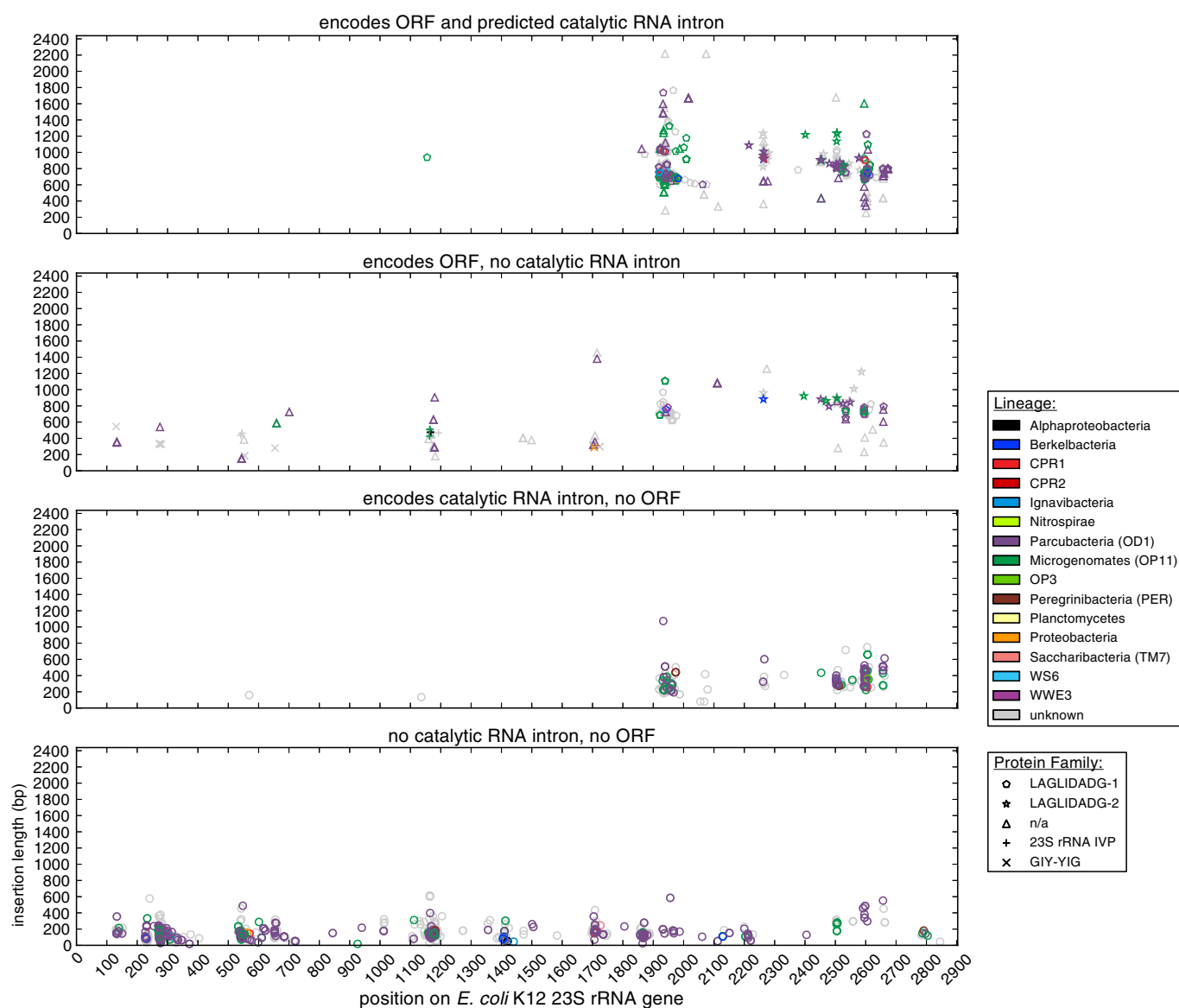
Extended Data Figure 6 | Features of insertion sequences encoded within 23S rRNA genes recovered from groundwater-associated bacteria. Bacteria associated with the CPR encode insertions within their 23S rRNA genes (Supplementary Table 5). These insertions share many features with those identified in 16S rRNA gene sequences from CPR bacteria. Taxonomy was determined by inclusion in a genome with an established phylogeny.
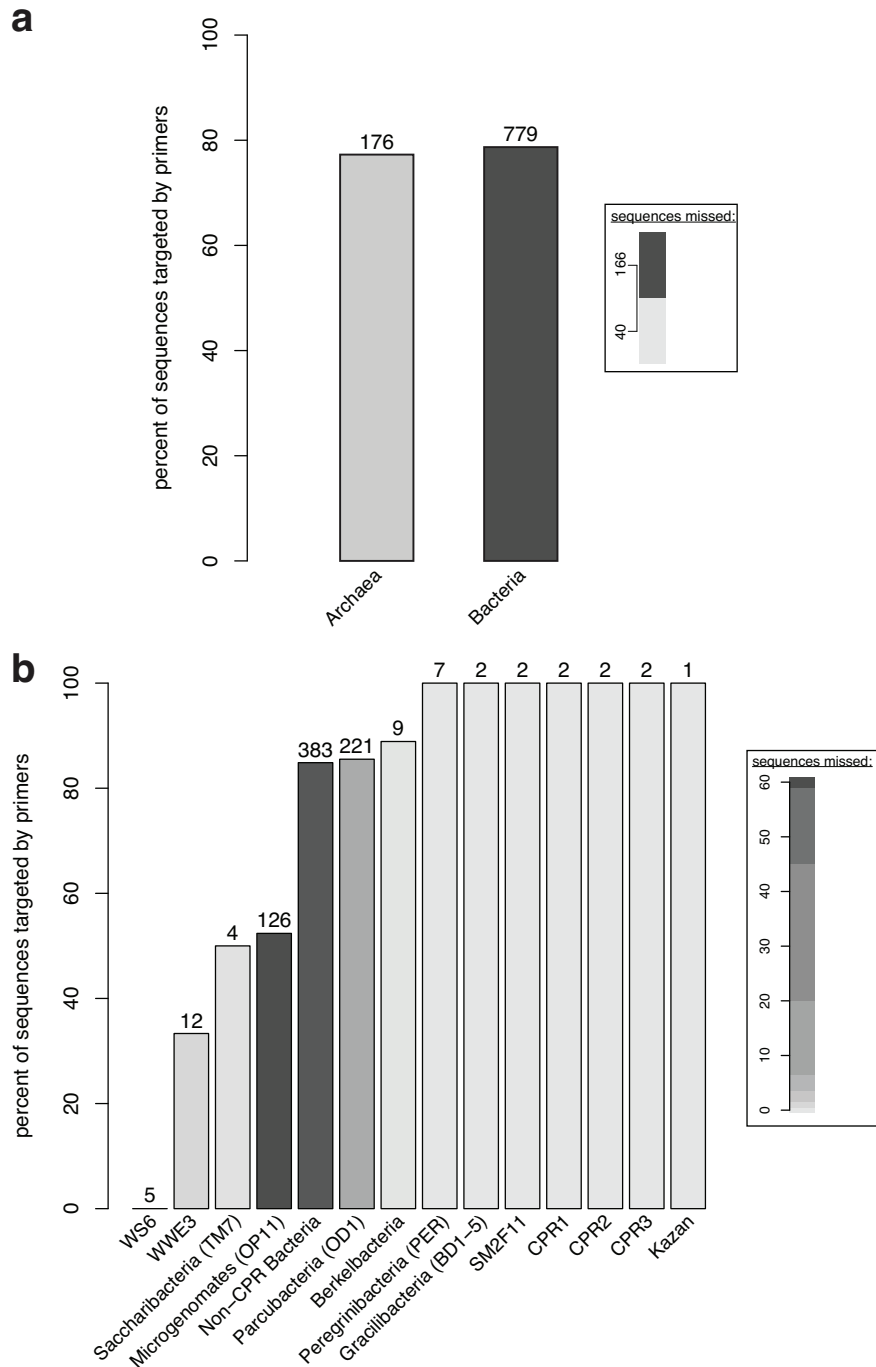
Extended Data Figure 7 | Analysis of the ability of PCR primers 515F and 806R to bind to recovered groundwater-associated 16S rRNA gene sequences. a, b, PrimerProspector was used to assess the ability of primers 515F and 806R to bind a non-redundant set of assembled near-complete 16S rRNA gene sequences (clustered at 97% sequence identity). The percentage of sequences that would be amplified by these primers is shown on the left axis, the total number of sequences analysed is on the top of each bar, and the number of sequences these primers would not bind to is indicated by the shading. Many assembled groundwater-associated 16S rRNA gene sequences would evade amplification by PCR primers 515F and 806R. Results of the analysis are shown at the domain (**a**) and superphylum or phylum (**b**) levels.
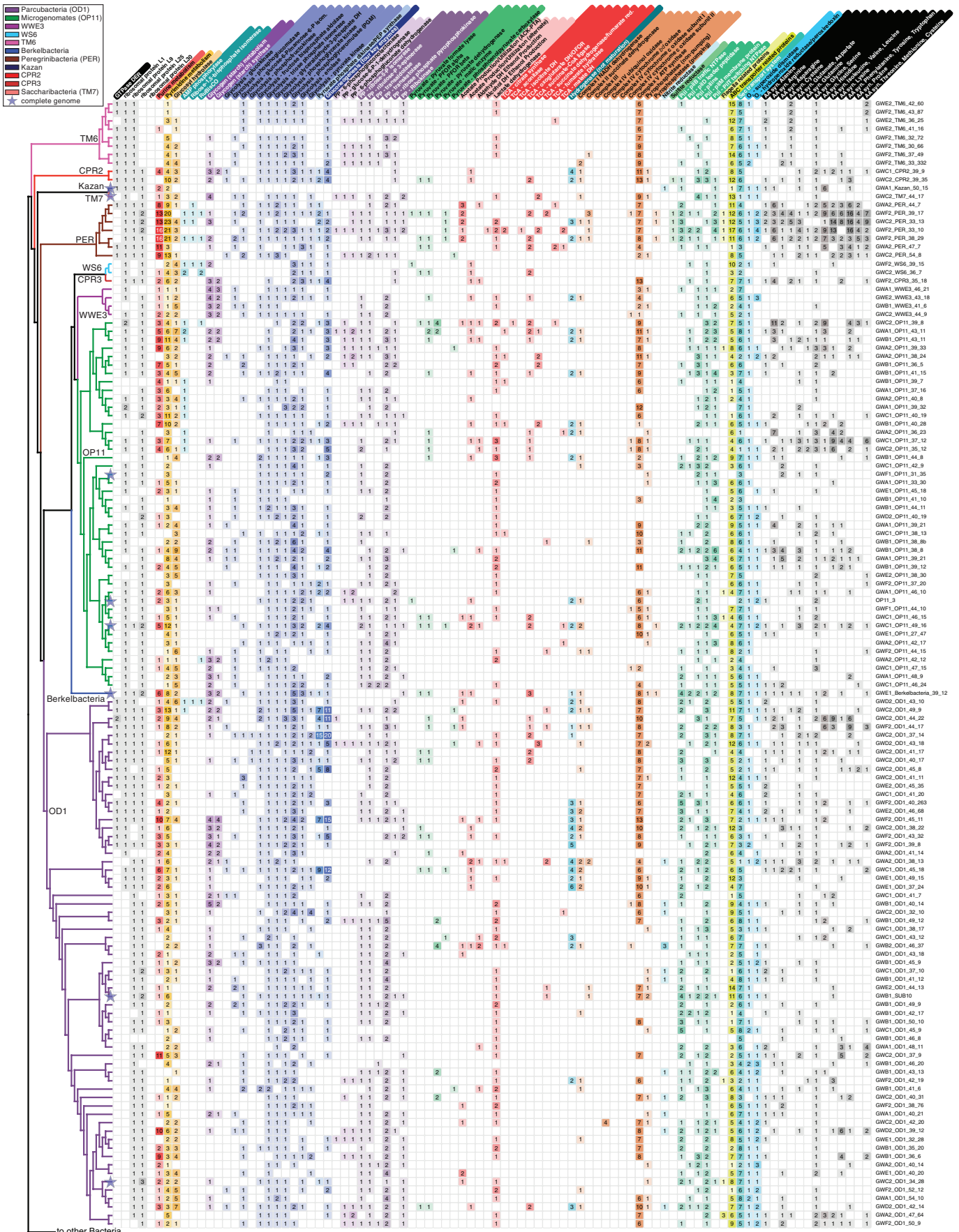
Extended Data Figure 8 | Metabolic potential and ribosomal protein analysis of genomes from CPR and TM6 organisms. Assembled genomes were analysed using ggKbase (Supplementary Data 4). Shown here is a non-redundant set of complete and near-complete genomes ($\geq$75% of single copy genes, $\leq$1.125 copies) organized based on a subset of a maximum-likelihood 16S rRNA gene phylogeny (Supplementary Fig. 1). CPR organisms have partial tricarboxylic acid (TCA) cycles and lack electron transport chain (ETC) complexes. In addition, they have incomplete biosynthetic pathways for nucleotides and amino acids. The Peregrinibacteria are a notable exception to some of these limitations. Several Parcubacteria exhibit a complete ubiquinol (cytochrome b_o) oxidase operon, as previously seen in Saccharibacteria[3]. However, lack of NADH dehydrogenase and other ETC components suggests that this enzyme is involved in oxygen scavenging/detoxification rather than energy production. AA Syn., amino acid synthesis; PP, pentose phosphate pathway.

Extended Data Table 1 | Proposed names for CPR phyla based on microbiology lifetime achievement award recipients

award	year awarded	recipient	proposed phylum name	superphylum
ASM Lifetime Achievement Award	2014	Roy Curtiss	Curtissbacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2013	Julian Davies	Daviesbacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2012	Stuart B. Levy	Levybacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2011	Susan Gottesman	Gottesmanbacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2010	Lucy Shapiro	Shapirobacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2009	Carl Woese	Woesebacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2008	Bernard Roizman	Roizmanbacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2007	Norman R. Pace	Pacebacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2006	R. John Collier	Collierbacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2005	Jonathan Beckwith	Beckwithbacteria	Microgenomates (OP11)
ASM Lifetime Achievement Award	2004	Alan Campbell	Campbellbacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	2003	Stanley Falkow	Falkowbacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	2002	Masayasu Nomura	Nomurabacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	2001	Bruce N. Ames	Amesbacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	2000	Boris Magasanik	Magasanikbacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	1999	Jonathan W. Uhr	Uhrbacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	1998	Charles Yanofsky	Yanofskybacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	1997	A. Dale Kaiser	Kaiserbacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	1996	Ralph S. Wolfe	Wolfebacteria	Parcubacteria (OD1)
ASM Lifetime Achievement Award	1995	Julius Adler	Adlerbacteria	Parcubacteria (OD1)
ISME Jim Tiedje Award	2014	Nancy Moran	Moranbacteria	Parcubacteria (OD1)
ISME Jim Tiedje Award	2012	Stephen Giovannoni	Giovannonibacteria	Parcubacteria (OD1)
ISME Jim Tiedje Award	2010	Bo Barker Jorgensen	Jorgensenbacteria	Parcubacteria (OD1)
ISME Jim Tiedje Award	2008	Norman R. Pace	Pacebacteria	Microgenomates (OP11)
ISME Jim Tiedje Award	2006	Gijs Kuenen	Kuenenbacteria	Parcubacteria (OD1)
ISME Jim Tiedje Award	2004	Farooq Azam	Azambacteria	Parcubacteria (OD1)

Human body epigenome maps reveal noncanonical DNA methylation variation

Matthew D. Schultz[1,2]†*, Yupeng He[1,2]*, John W. Whitaker[3]†, Manoj Hariharan[2], Eran A. Mukamel[4,5], Danny Leung[6], Nisha Rajagopal[6], Joseph R. Nery[2], Mark A. Urich[2], Huaming Chen[2], Shin Lin[7], Yiing Lin[8], Inkyung Jung[6], Anthony D. Schmitt[6], Siddarth Selvaraj[1], Bing Ren[6,9], Terrence J. Sejnowski[4,10,11], Wei Wang[3,12] & Joseph R. Ecker[2,11]

Understanding the diversity of human tissues is fundamental to disease and requires linking genetic information, which is identical in most of an individual's cells, with epigenetic mechanisms that could have tissue-specific roles. Surveys of DNA methylation in human tissues have established a complex landscape including both tissue-specific and invariant methylation patterns[1,2]. Here we report high coverage methylomes that catalogue cytosine methylation in all contexts for the major human organ systems, integrated with matched transcriptomes and genomic sequence. By combining these diverse data types with each individuals' phased genome[3], we identified widespread tissue-specific differential CG methylation (mCG), partially methylated domains, allele-specific methylation and transcription, and the unexpected presence of non-CG methylation (mCH) in almost all human tissues. mCH correlated with tissue-specific functions, and using this mark, we made novel predictions of genes that escape X-chromosome inactivation in specific tissues. Overall, DNA methylation in several genomic contexts varies substantially among human tissues.

To understand the variability of DNA methylation across human tissues better, we obtained post-mortem samples of 18 tissue types from 4 individuals (5 singletons, 8 duplicates and 5 triplicates; Fig. 1a, Supplementary Methods and Supplementary Table 1) and performed deep transcriptome (36 messenger-RNA-seq samples; 120–475 million reads per sample), base-resolution methylome (36 MethylC-seq[4] samples; 30–80× genome coverage per sample), and genome sequencing (4 whole genome sequences; 20–45× genome coverage per sample). We focused our initial analysis on cytosines in the CG context and used a previously published method[2] to identify differential methylation (Supplementary Methods). We found that 15.4% (4,073,896 out of 26,474,560 sites tested) of CG sites in these experiments are strongly differentially methylated (minimum methylation difference ≥ 0.3; Extended Data Fig. 1a), which is similar to a previous study[2]. To identify differentially methylated regions (DMRs), we combined sites within 500 base pairs (bp) of one another and found 1,198,132 DMRs. Even with these stringent criteria, 719,837 (60.1%) of the DMRs we identified were novel[2,5].

As expected, hypomethylation at DMRs correlated with tissue-specific functions[2,6]. For example, strongly hypomethylated DMRs in the aorta overlap with aorta-specific super enhancers[7] around *MYH10*, a gene involved in blood vessel function[8] (Fig. 1b). To validate our DMRs further, we performed hierarchical clustering on their weighted methylation levels[9] (Supplementary Methods, Fig. 1c and Extended Data Fig. 1b, c). Tissues that were part of the same organ system

clustered together (for example, heart and muscle tissues). We compared these results to a clustering of differentially expressed genes identified in the transcriptomes and found a similar separation of organ systems (Supplementary Methods, Fig. 1d and Extended Data Fig. 1d). Furthermore, Genomic Regions Enrichment of Annotations Tool[10] analysis on the most hypomethylated tissue-specific DMRs revealed many tissue-specific functions (Extended Data Fig. 1e, f, Supplementary Methods and Supplementary Tables 2–3).

To examine the relationship between methylation and transcription, we correlated the methylation levels of DMRs and the expression of the closest genes (Fig. 2a, Extended Data Fig. 2a, b and Supplementary Methods). As expected, methylation in DMRs had a negative correlation with expression, and this correlation grew stronger closer to the transcription start site. The strongest negative correlation was not in gene promoters but downstream of the promoter up to 8 kilobases (kb) away (intragenic (0.3 kb to 8 kb) versus promoter region and upstream region (-2 kb to 0.3 kb) median Spearman correlation coefficient difference -0.07; Mann–Whitney $P = 4.2 \times 10^{-17}$; Fig. 2a). This analysis shows that transcription is strongly associated with intragenic DMRs in the tissues we examined, extending similar observations in cancer methylomes[11].

These intragenic methylation differences have previously been suggested to mark intragenic CG islands (CGIs) or CGI shores[5,12–14]. However, only a small fraction of intragenic DMRs fell in these features (19%; Extended Data Fig. 2c). In addition, predicted enhancers and putative promoters only accounted for 23% and 22% of intragenic DMRs, respectively, suggesting that the remaining DMRs, which we call undefined intragenic DMRs (uiDMRs), represent an unrecognized set of functional elements (35%; Extended Data Fig. 2c and Supplementary Methods). The methylation level of these uiDMRs correlated strongly with the expression of the genes containing them. To examine their regulatory potential, we plotted their histone modification profiles (histone 3 Lys 4 methylation (H3K4me1), H3K4me3, H3K27ac, H3K9me3, H3k27me3 and H3K36me3) derived from the same tissue samples[15] and found five classes: weak enhancer, promoter-proximal, transcribed, poised enhancer and unmarked (Extended Data Figs 2d–h, 3a, b and Supplementary Methods). Classes with strong, active histone modifications were moderately negatively correlated with expression (weak enhancer and proximal promoter uiDMRs; median Spearman correlation coefficient -0.32 and -0.16, respectively); whereas, uiDMRs with less active histone modifications exhibited a weak negative correlation (transcribed and poised enhancer uiDMRs). Notably, the correlation between expression and methylation at promoter-proximal

Figure 1 | The methylomes and transcriptomes of human tissues. a, The tissues analysed in this study. Samples are denoted by the two letter code in parentheses followed by an individual ID. **b,** Browser screenshot of an example DMR. The top track contains gene models. The following four tracks contain green blocks indicating the location of super enhancers, enhancers and hypomethylated DMRs in the aorta, respectively. The remaining tracks display methylation data from each sample. Gold ticks are CG sites with heights proportional to their methylation level. Ticks on the forward and reverse strand are projected upward and downward from the dotted line, respectively. **c, d,** Hierarchical clustering of DMR methylation levels (**c**) and expression levels of differentially expressed genes (**d**). Colours indicate the organ systems each sample belongs to.

uiDMRs was as strong as the correlation with intragenic DMRs that overlapped strong promoters (Extended Data Fig. 4 and Supplementary Methods), indicating that intragenic promoter and promoter-proximal sequences are more predictive of changes in methylation than those enriched for enhancer-like chromatin modifications.

By contrast, unmarked uiDMRs showed a weakly positive correlation with expression (Extended Data Fig. 4d). Notably, we found many of the motifs enriched in tissue-specific uiDMRs were present in tissue-specific enhancers (for example, HNF4a (ref. 16) in liver-specific uiDMRs), suggesting that these DMRs are tissue-specific regulatory elements (Supplementary Methods and Supplementary Tables 4 and 5). Recently, hypomethylated regions that appear inactive in adult tissues but active during fetal development were identified in mice[6]. We examined the DNase I hypersensitivity profiles of unmarked uiDMRs in matched fetal tissues[17] and found an enrichment of hypersensitivity (Extended Data Fig. 5 and Supplementary Table 6), suggesting that hypomethylation of inactive DMRs can be maintained at regions active earlier in development.

We next examined whether variation in methylation is associated with genetic variation across individuals, which has not been widely characterized in healthy primary tissues or using whole-genome bisulphite sequencing[18,19]. To identify individual-specific DMRs, we used a method[20] that is sensitive to these differences unlike the methodology used above (Supplementary Methods). We first restricted our analysis to triplicated samples and ranked DMRs by a tissue-specific methylation outlier score that is largest when the methylation level in one individual differs from the other two. We found an ~1.6-fold enrichment of single nucleotide polymorphisms (SNPs) associating with methylation changes in the top 2,500 methylation-outlier-score-ranked DMRs in all tissues (Supplementary Methods). We then used the Epigram pipeline[21] to predict tissue-specific methylation from DNA motifs in these DMRs and found them highly predictive (average area under the curve (AUC) 0.79; Supplementary Methods). These full models used an average of 156 motifs; however, an average AUC of 0.74 was achieved using only 20 core transcription factor motifs per tissue.

We then identified groups of corresponding motifs by clustering the sets of tissue-specific motifs (Supplementary Methods). The motif groups were clustered by their tissue hypo- and hypermethylation specificities (Fig. 2b). In total, 42 out of 95 motifs only had hypomethylation specificity; for example, *MEIS*, which is involved in heart development[22], is hypomethylated in the left ventricle, right atrium and right ventricle. We also identified 34 motifs enriched at both hypomethylated DMRs in some tissues, and in hyper-methylated DMRs in some other tissues. Three of these motifs match transcription factor families (*FOX*, *HOX* and *GATA*) and are most significantly enriched in hypomethylated regions, suggesting that they are primarily involved in regulating hypomethylation.

Mammalian cells have high genome-wide levels of mCG, with the exception of a cultured human fetal fibroblast cell line (IMR90)[4], cancer cells[23,24] and placenta (PLA)[25]. Surprisingly, large regions of the pancreatic methylomes (PA-2 and PA-3) were significantly hypomethylated (Extended Data Fig. 6a). We developed a method to identify partially methylated domains (PMDs) genome-wide (Supplementary Tables 7–8 and Supplementary Methods) and found pancreatic PMDs were smaller than those in IMR90 and PLA (Extended Data Fig. 6b) and covered a smaller fraction of the genome (Fig. 2c). All pairs of PMDs overlapped significantly, indicating that these regions are largely shared (>40% overlap; $P < 0.001$; Extended Data Fig. 6c).

Genes in samples with PMDs are transcriptionally repressed[25,26], but these regions also show reduced expression in all of the tissues we surveyed whether or not a PMD is present (Fig. 2d). In both IMR90 and PA-2, these regions showed an enrichment in repressive modifications (H3K27me3 and H3K9me3; median difference 0.025–0.168 reads per kilobase per million (RPKM); Mann–Whitney $P < 2.51 \times 10^{-161}$) and a depletion in active modifications (H3K4me1, H3K27ac and H3K36me3; median difference 0.050–0.012 RPKM; Mann–Whitney $P < 2.03 \times 10^{-53}$) compared to shuffled regions (Fig. 2e, f, Extended Data Fig. 6 d, e and Supplementary Methods), which provides a potential mechanism for their repression. To try to account for this global hypomethylation, we plotted the expression levels of *DNMT1*, *DNMT3A*, *DNMT3B* and *DNMT3L* but found no
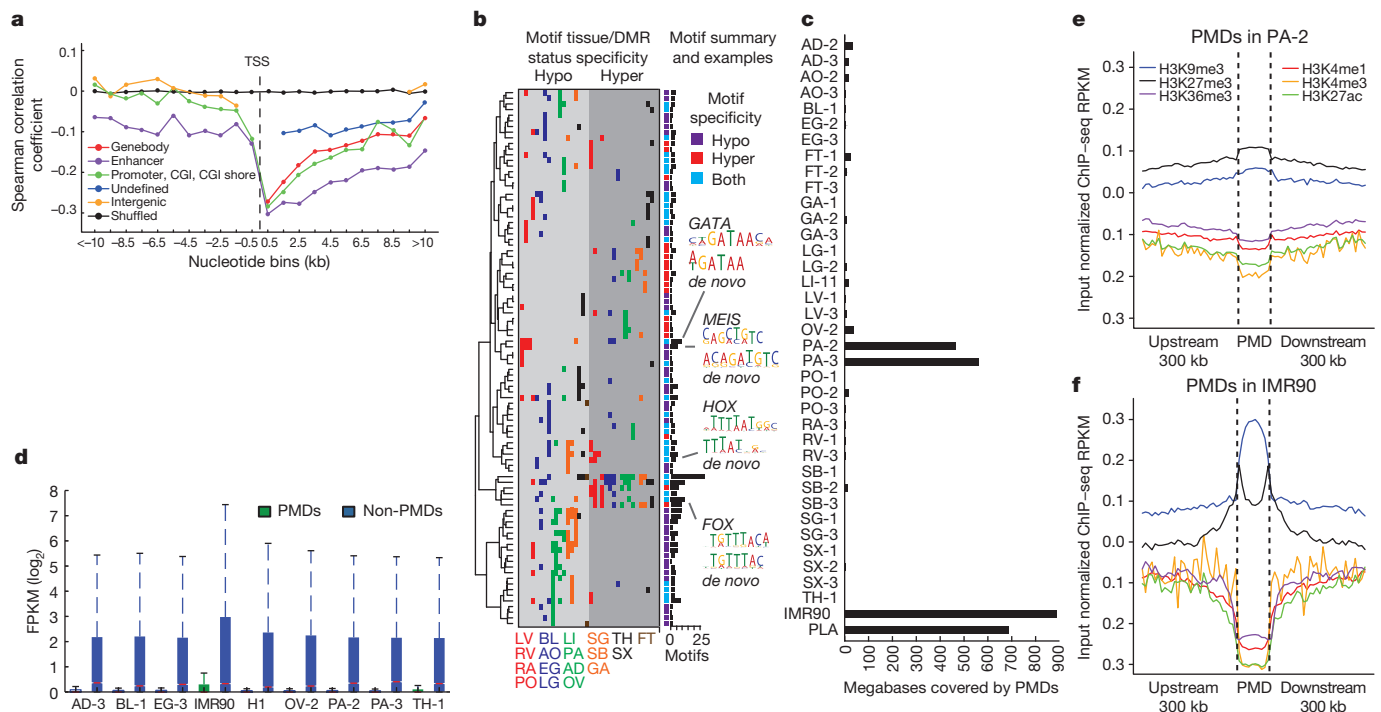
Figure 2 | DNA methylation and its relationship with gene expression.
a, The mean Spearman correlation coefficient at various distances between the methylation level of autosomal DMRs and the expression of the nearest gene. These correlations are shown for DMRs: overlapping genes (gene body), overlapping enhancers, overlapping promoters or CpG islands (CGIs) or CGI shores, not overlapping genes (intergenic) and all remaining DMRs (undefined). TSS, transcription start site. **b**, Heatmap showing the tissue-specific methylation preference of each motif. The tissues are coloured according to Fig. 1c, and the ordering is listed at the bottom of the figure.

The bar plot on the right shows the number of times the motif was present in the 20 motif models. **c**, The number of base pairs covered by PMDs in all samples. **d**, The distribution of expression inside and outside of PA-2 PMDs across various samples. Notches indicate a confidence interval estimated from 1,000 bootstrap samples. Each PMD boxplot consists of 3,627 genes, and each non-PMD boxplot consists of 22,907 genes. FPKM, fragments per kilobase of transcript per million mapped reads. **e**, **f**, Histone modification profiles in and around PMDs in PA-2 (**e**) and IMR90 (**f**).

systematic expression difference between samples with and without PMDs (Extended Data Fig. 7a–d).

Previous studies have highlighted the existence of methylation outside of the CG context (mCH) in human embryonic stem cells[4], brain[1,20] and at the promoter of the *PGC-1α* gene (*PPARGC1A*) in skeletal muscle[27]. We found evidence for appreciable amounts of mCH in many of these tissues (Fig. 3a and Extended Data Fig. 8a). A 5-bp motif split the samples into two groups, one with mCH enriched in a TNCAC motif and another with mCH enriched in an NNCAN motif (where N is any base) (Supplementary Methods). The TNCAC motif is highly similar to the one previously identified in purified glia (GLA) and neurons (NRN) (TACAC). These motifs differ from those found in H1 embryonic stem cells (H1) and induced pluripotent stem cells (TACAG)[4,26] (Fig. 3b–d). We quantified the extent of mCH across these samples by plotting the distribution of methylation levels at mCH sites in the 25 samples with a TNCAC motif, which revealed a methylation level similar to that of GLA, NRN and H1 (Extended Data Fig. 8b)[4,20]. Most of the tissue types were consistently enriched for the TNCAC or NNCAN motif, but several (oesophagus, lung, pancreas and spleen) had replicates that disagreed, suggesting that mCH is not homogenously distributed across these tissues.

To examine the potential functional effect of mCH in adult tissues, we plotted the distribution of expression levels for various quantiles of gene body mCH as it was previously reported to be positively correlated with expression in H1 (ref. 4) and negatively correlated with expression in neurons[20]. This analysis revealed a negative correlation between expression and mCH (Extended Data Fig. 8c and Supplementary Methods). Next, we combined our replicates and clus-

tered genes by the patterns of CAS methylation (in which S is a G or C) in and around their gene body (Fig. 3e and Supplementary Methods). To characterize the genes assigned to each cluster, we performed DAVID functional annotation clustering (Supplementary Table 9 and Supplementary Methods), which revealed several different classes. Clusters 1, 2, 16 and 19 contained genes highly enriched for terms involved in basic cellular processes and had an active methylation state (that is, hypermethylation in embryonic samples and hypomethylation in tissue and brain samples) across all samples. Clusters 5 and 6 were dominated by terms related to neuronal function and genes in this class were differentially methylated between neurons and glia and have inactive methylation states in other samples (that is, hypomethylation in embryonic samples and hypermethylation in tissue and brain samples). Cluster 12 was enriched for heart- and muscle-related terms and its genes had an active methylation state in the three heart tissues as well as a weakly active methylation state in psoas but appeared inactive in other samples. Lastly, cluster 14 possessed an active methylation state in brain and tissue samples but was inactive in embryonic samples. Despite being inactive in the H1 samples, this class of genes was highly enriched for terms related to development.

To define the transition of mCH motifs over development better, we examined the ratio of the methylation level of CAC and CAG (mCAC and mCAG) sites in a variety of differentiated (tissues, NRN and GLA), embryonic (H1), and embryonic-derived (neural progenitor cells (NPC), mesendoderm (MES), trophoblast-like (TRO), mesenchymal stem cells (MSC))[28] cell samples (Fig. 3f). With the exception of brain cells, mCH levels drop during differentiation, and the mCAC/mCAG ratios revealed a shift in motif usage across developmental time (Fig. 3f); although, mCAC and mCAG within the same gene remain
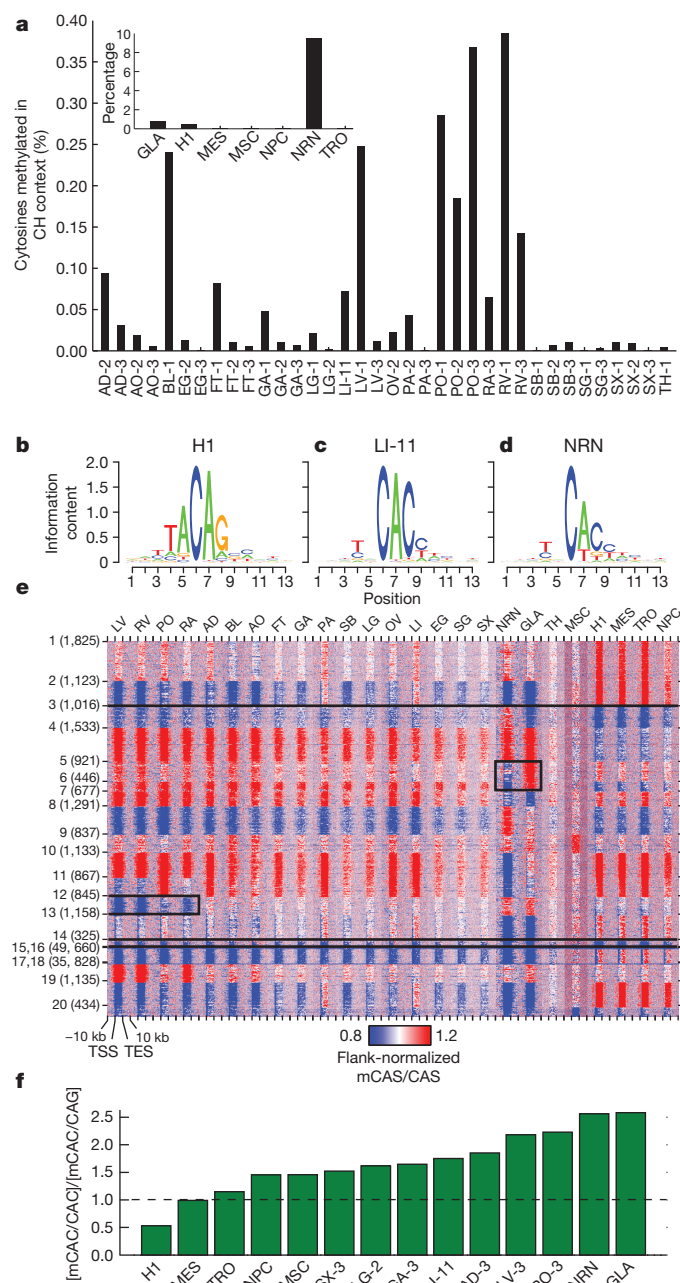
Figure 3 | mCH is prevalent in human tissues. a, The fraction of methylated cytosines in the CH context by sample. **b–d**, Representative mCH motifs from embryonic (H1; **b**), tissue (LI-11; **c**) and brain (NRN; **d**) samples. The height of each letter represents its information content. **e**, Heatmap of genic mCAS patterns normalized to the flanking region. Each gene was assigned to 1 of 20 clusters, which is indicated by the number and tick marks on the y axis. The tick marks on the x axis indicate the upstream, transcription start, transcription end, and downstream segments of each gene. The boxes around various patterns highlight regions referenced in the main text. TES, transcription end site. **f**, Bar plot of the ratio of the genome-wide mCAC to mCAG in various samples.

tightly correlated in both early embryonic and differentiated tissues (Extended Data Fig. 8d, e).

Methylation has previously been shown to be predictive of genes escaping X-chromosome inactivation in neurons[20]. We investigated this phenomenon in these samples by comparing the promoter mCG and gene body mCH of genes that had previously been identified to escape X-chromosome inactivation[29] in 11 tissues with mCH (Fig. 4a). Female-specific promoter mCG hypomethylation and gene body mCH hypermethylation were present at escapee genes at a similar
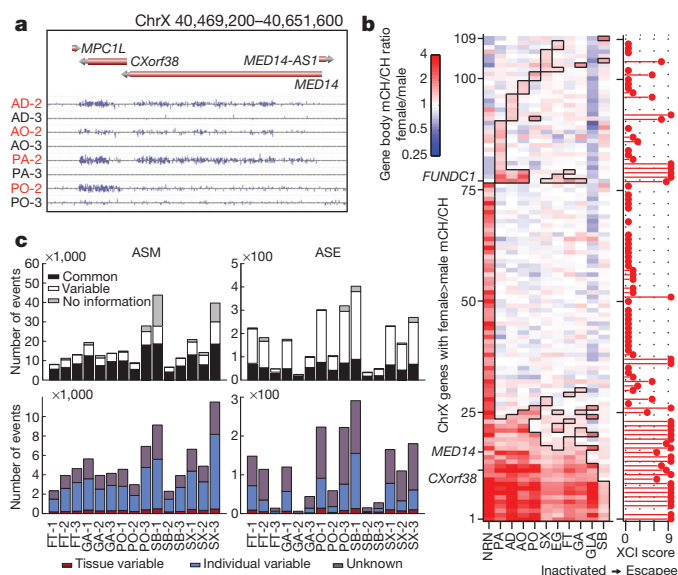
Figure 4 | Allele-specific methylation and expression. a, Browser screenshot of the increase in female mCH for a gene known to escape X-chromosome inactivation (*MED14*). Sample names are coloured by gender (male, black; female, red). *MED14-AS1* is also known as *MED14OS*. **b**, Ratio of mCH level in female versus male samples across genes with a significant difference in at least one sample. Cells boxed in black denote samples with a statistically significant difference between females and males. The XCI score for each gene is from ref. 29 and indicates the degree of escaping X-chromosome inactivation. **c**, The number of ASM and ASE sites across the triplicated tissues. The top row depicts ASM events (left) and ASE events (right) that are allele-specific in all tissues (black), are variable across tissues (white), or do not possess enough data to tell (grey). The bottom row depicts the distribution of variable sites from the top row that vary by individual (blue), tissue (red) or neither (purple).

level as in neurons[20] (Extended Data Fig. 9a). Using these tissue methylomes, gene body mCH was appreciably predictive of biallelically expressed genes (AUC 0.89; Extended Data Fig. 9b and Supplementary Methods). To a lesser extent, we observed female-specific promoter mCH and gene body mCG hypermethylation at escapee genes (Extended Data Fig. 9a, c, d). Although female-specific promoter mCG hypomethylation, promoter mCH hypermethylation and gene body mCG hypermethylation are predictive of X-chromosome inactivation escapees, female-specific gene body mCH hypermethylation is the most predictive feature of X-chromosome inactivation escapees (Extended Data Fig. 9a, b–e). We detected female-specific mCH hypermethylation in 109 out of 612 X-linked genes, including 9 genes hypermethylated in all 11 tissues and 72 genes that were hypermethylated in only one tissue (Fig. 4b). Several genes such as *FUNDC1* showed female-specific hypermethylation in several tissues but not in neurons, suggesting a tissue-dependent regulation of the escape from X inactivation.

Allele-specific methylation and expression (ASM and ASE, respectively) may also have a role in the regulation of autosomal genes. To examine these phenomena in human tissues, we combined the RNA-seq and MethylC-seq data sets with phased genotypes for each individual in this study[3,15] (Extended Data Fig. 10a and Supplementary Methods). Using the triplicate tissue samples (fat (FT), gastric (GA), psoas (PO), small bowel (SB) and spleen (SX)), we identified 8,464–48,560 ASM events in the CG context and 48–403 ASE genes across these tissues (Supplementary Tables 10, 11 and Supplementary Methods). We next looked for ASM events that varied across individuals within a tissue-type (tissue variable) and those that varied across a tissue-type within an individual (individual variable). Of the ASM events that varied, 4.1–7.5% and 54.5–70.0% were individual- and tissue-variable, respectively; whereas, of the ASE events that varied, 0.0–20.0% were individual-variable and 13.3–48.8% were

tissue-variable (Fig. 4c and Supplementary Methods). Of the ASE events, 38.4–87.4% had an ASM event within 100 kb, and of these sites, 76% had an ASM and ASE event that was matched (that is, a DMR was hypomethylated on the same haplotype as the more highly expressed allele). Furthermore, we found that a larger fraction of ASE genes were observed near ASM events whether or not the events matched (Extended Data Fig. 10 b, c and Supplementary Methods). These results demonstrate a link between ASM and ASE in human tissues.

Here we have presented the deepest set of base resolution maps of mCG and mCH so far along with chromatin modification states, haplotype-resolved genome sequences and transcriptional profiles for a large set of human tissues. These data sets allowed us to identify *cis*-regulatory elements accurately. Furthermore, they revealed the existence of mCH genome-wide in a subpopulation of cells from differentiated human tissues, which seems to be repressive. Our analysis of genic mCH across human tissues indicates a tissue-specific distribution that is distinct from those genes that were previously identified in embryonic stem cells and the brain. These genes are enriched for a variety of functions, most surprisingly those involved in development. These analyses raise the intriguing possibility that mCH is used in adult stem cells[30] and could help to repress these genes as the cells transition into their differentiated role.

1. Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23,** 555–567 (2013).
2. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–481 (2013).
3. Selvaraj, S., Dixon, J. R., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nature Biotechnol.* **31,** 1111–1118 (2013).
4. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
5. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41,** 178–186 (2009).
6. Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genet.* **45,** 1198–1206 (2013).
7. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
8. Yuen, S. L., Ogut, O. & Brozovich, F. V. Nonmuscle myosin is regulated during smooth muscle contraction. *Am. J. Physiol. Heart Circ. Physiol.* **297,** H191–H199 (2009).
9. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* **28,** 583–585 (2012).
10. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* **28,** 495–501 (2010).
11. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510,** 537–541 (2014).
12. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–257 (2010).
13. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41,** 1350–1353 (2009).
14. Deaton, A. M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* **21,** 1074–1086 (2011).
15. Leung, D. *et al.* Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518,** 350–354 (2015).
16. Parviz, F. *et al.* Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nature Genet.* **34,** 292–296 (2003).
17. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).
18. Gutierrez-Arcelus, M. *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* **2,** e00523 (2013).
19. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnol.* **31,** 142–147 (2013).
20. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341,** 6146 (2013).
21. Whitaker, J. W., Chen, Z. & Wang, W. Predicting the human epigenome from DNA motifs. *Nature Methods* **12,** 265–272 (2015).
22. Stankunas, K. *et al.* Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ. Res.* **103,** 702–709 (2008).
23. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22,** 246–258 (2012).
24. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature Genet.* **44,** 40–46 (2011).
25. Schroeder, D. I. *et al.* The human placenta methylome. *Proc. Natl Acad. Sci. USA* **110,** 6037–6042 (2013).
26. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
27. Barrès, R. *et al.* Non-CpG methylation of the PGC-1α promoter through DNMT3B controls mitochondrial density. *Cell Metab.* **10,** 189–198 (2009).
28. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).
29. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434,** 400–404 (2005).
30. Wagers, A. J. & Weissman, I. L. Plasticity of adult stem cells. *Cell* **116,** 639–648 (2004).

Author Contributions B.R., T.J.S., W.W. and J.R.E. designed and supervised research. S.L. and Y.L. collected tissues. J.R.N. and M.A.U. conducted MethylC-seq, RNA-seq and genome sequencing experiments. D.L. conducted ChIP-seq experiments. N.R. performed ChIP-seq data analysis. M.D.S., Y.H., M.H. and H.C. performed sequencing data processing. J.W.W. performed motif prediction and mutation analysis. M.D.S. designed and implemented the methylation processing and analysis module. M.D.S., Y.H., J.W.W., M.H. and E.A.M. performed statistical and bioinformatic analyses. M.D.S., Y.H., J.W.W. and J.R.E. prepared the manuscript.

Extended Data Figure 1 | Identification of DMRs and multidimensional scaling analysis. a, Line plot showing the fraction of differentially methylated CG sites (dynamic CGs) out of all CG sites under various methylation difference cutoffs. The methylation difference of a CG site is defined in ref. 2. The grey line indicates the cutoff (0.3) used to call differentially methylated sites. **b,** A plot of the first two principal components from the methylation level multidimensional scaling. Tissues are shaded by the organ group they belong to as in Fig. 1c, d. **c, d,** Bar charts of the cumulative amount of variance explained by the first *N* principal components from the multidimensional scaling performed on the methylation levels of all DMRs (**c**) and the expression levels of all differentially expressed genes (**d**). **e,** A representative example of enriched Gene Ontology biological process terms based on the most hypomethylated DMRs from LV-1. **f,** A representative example of enriched mouse phenotype terms based on the most hypomethylated DMRs from LV-1.

Extended Data Figure 2 | DMRs and their correlation with transcription.
a, A browser screenshot of an example DMR downstream of the transcription start site. **b**, Expression level of the *BIN1* gene that contains the DMR in **a**. **c**, The percentages of hypomethylated intragenic DMRs in each class of genomic features. **d–h**, Histone modification profiles of five categories of uiDMRs.
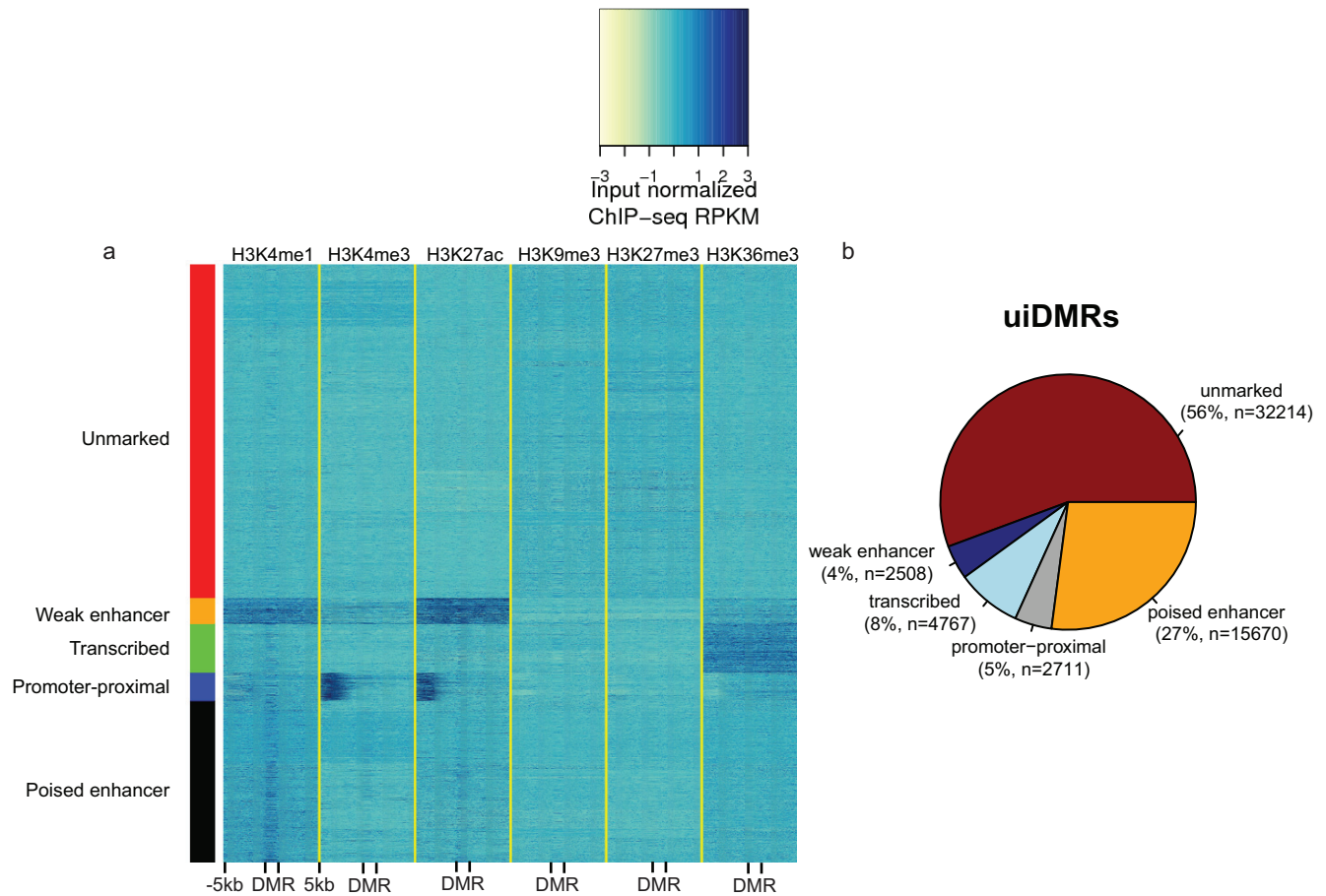
Extended Data Figure 3 | Classification of uiDMR histone profiles and uiDMR properties. a, Heatmap of the histone modification profiles for the five types of uiDMRs. The profiles were plotted for each mark across the DMR and the 5 kb upstream and downstream and the colours of each cell indicate the input normalized ChIP-seq RPKM. The colours on the left indicate the group of each profile assigned by *k*-means clustering (red, weak enhancer; orange, promoter-proximal; green, transcribed; blue, unmarked; black poised enhancer). **b**, A pie chart of the distribution of uiDMRs across the classes defined by *k*-means clustering.
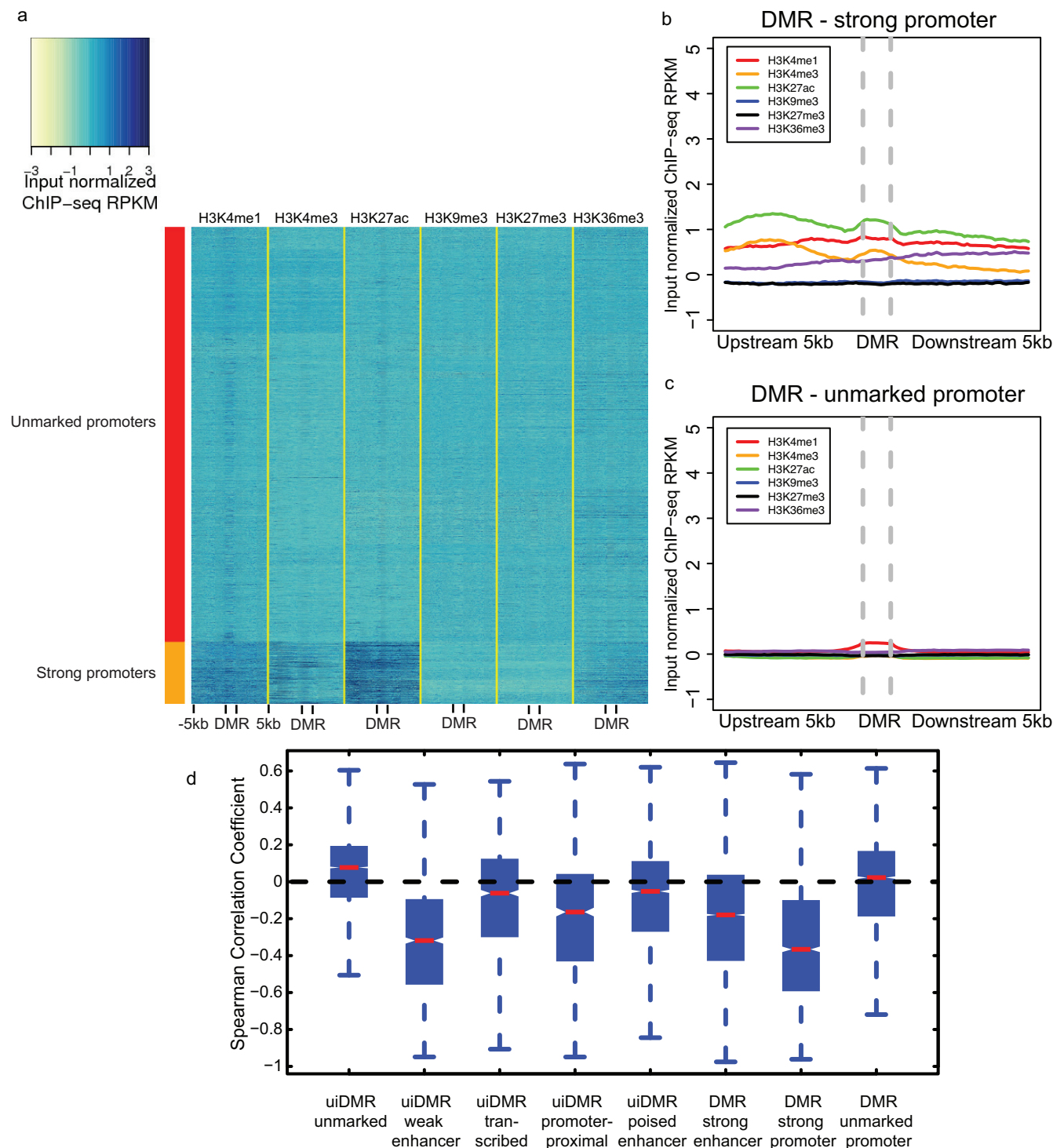
Extended Data Figure 4 | Classification of promoter histone profiles. a, A heatmap of the histone modification profiles across strong (rows labelled with red) and unmarked (rows labelled with orange) promoters. The profiles were plotted for each mark across the promoter and the 5-kb upstream and downstream, and the colours of each cell indicate the input normalized ChIP-seq RPKM. **b, c,** The aggregate profiles for strong (**b**) and unmarked (**c**) promoters, respectively. **d,** The distribution of the Spearman correlation coefficients between the methylation level of different types of hypomethylated intragenic DMRs and the expression of the nearest gene. Notches indicate a confidence interval estimated from 1,000 bootstrap samples.
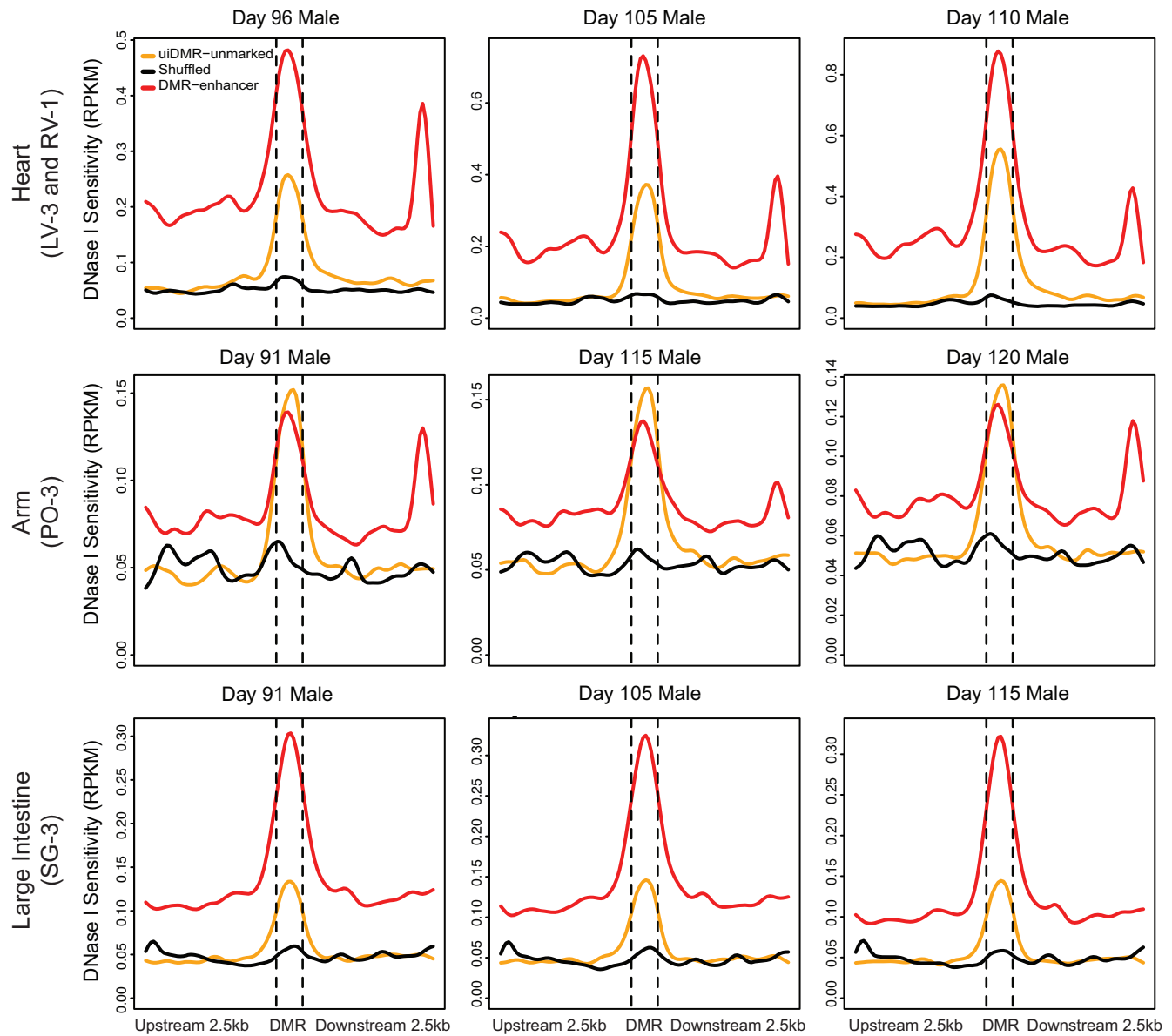
Extended Data Figure 5 | uiDMR fetal DNase I profiles. DNase I profiles of various fetal tissues corresponding to the tissues presented in this study. The samples are arranged column-wise by age, and row-wise by fetal tissue. The uiDMR-unmarked line represents the DNase I profile of uiDMRs without histone modifications. The DMR-enhancer line represents the DNase I profile of DMRs that overlapped an enhancer in a matched tissue in this study (indicated in the row label in parentheses). The shuffled line represents the DNase I profile of uiDMRs randomly shuffled across the genome.
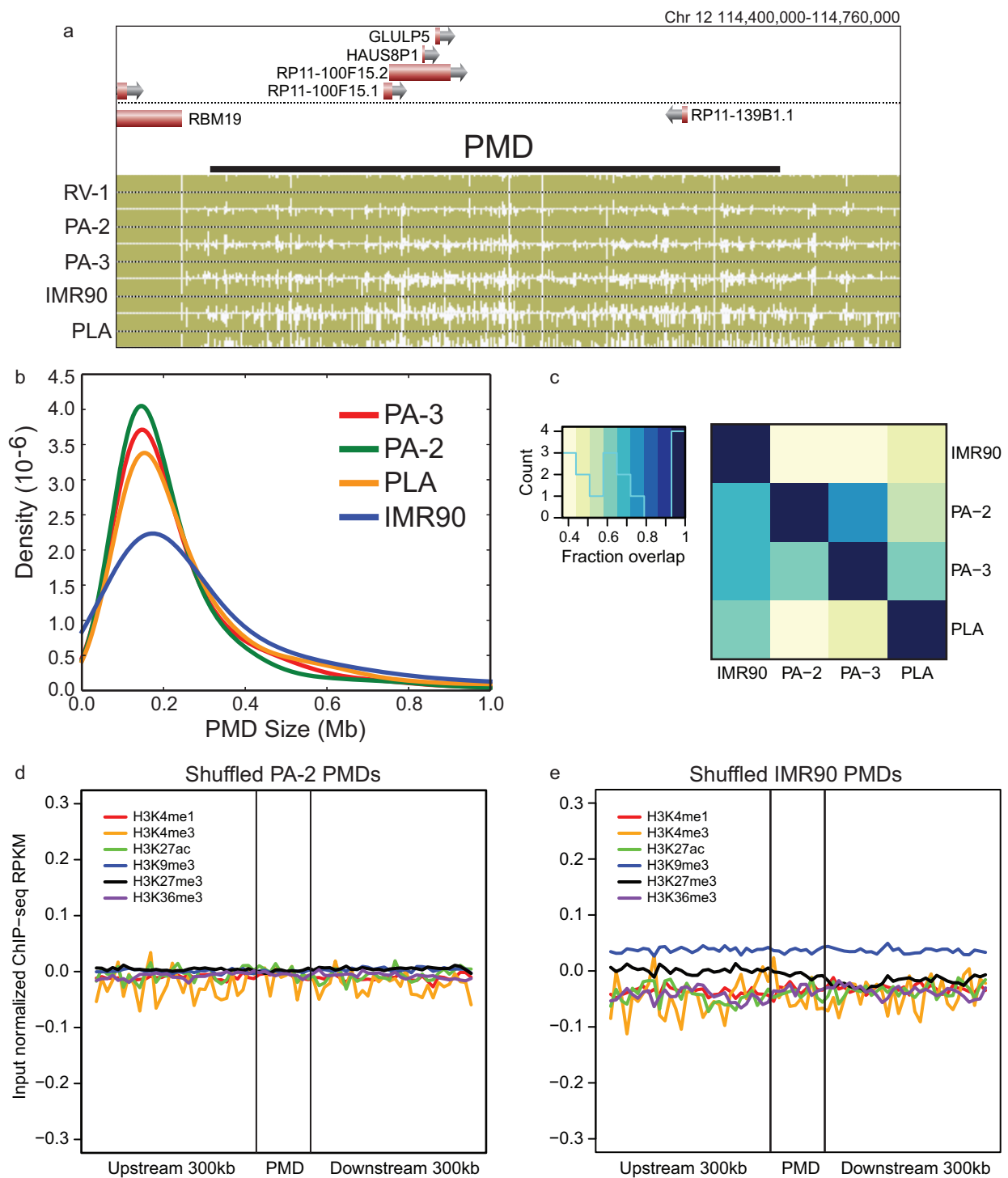
Extended Data Figure 6 | PMD features. a, A browser screenshot (see Fig. 1 for description) of an example PMD found in IMR90, PLA, PA-2 and PA-3. RV-1 is included as a representative sample without PMDs. **b,** The distribution of sizes of PMDs in various samples. **c,** A heatmap representation of the overlap between various sets of PMDs. The denominator of the fraction of overlap is determined by the sample on the *y* axis. **d, e,** ChIP-seq profiles of the PMD regions defined in PA-2 (**c**) and IMR90 (**d**) after shuffling.
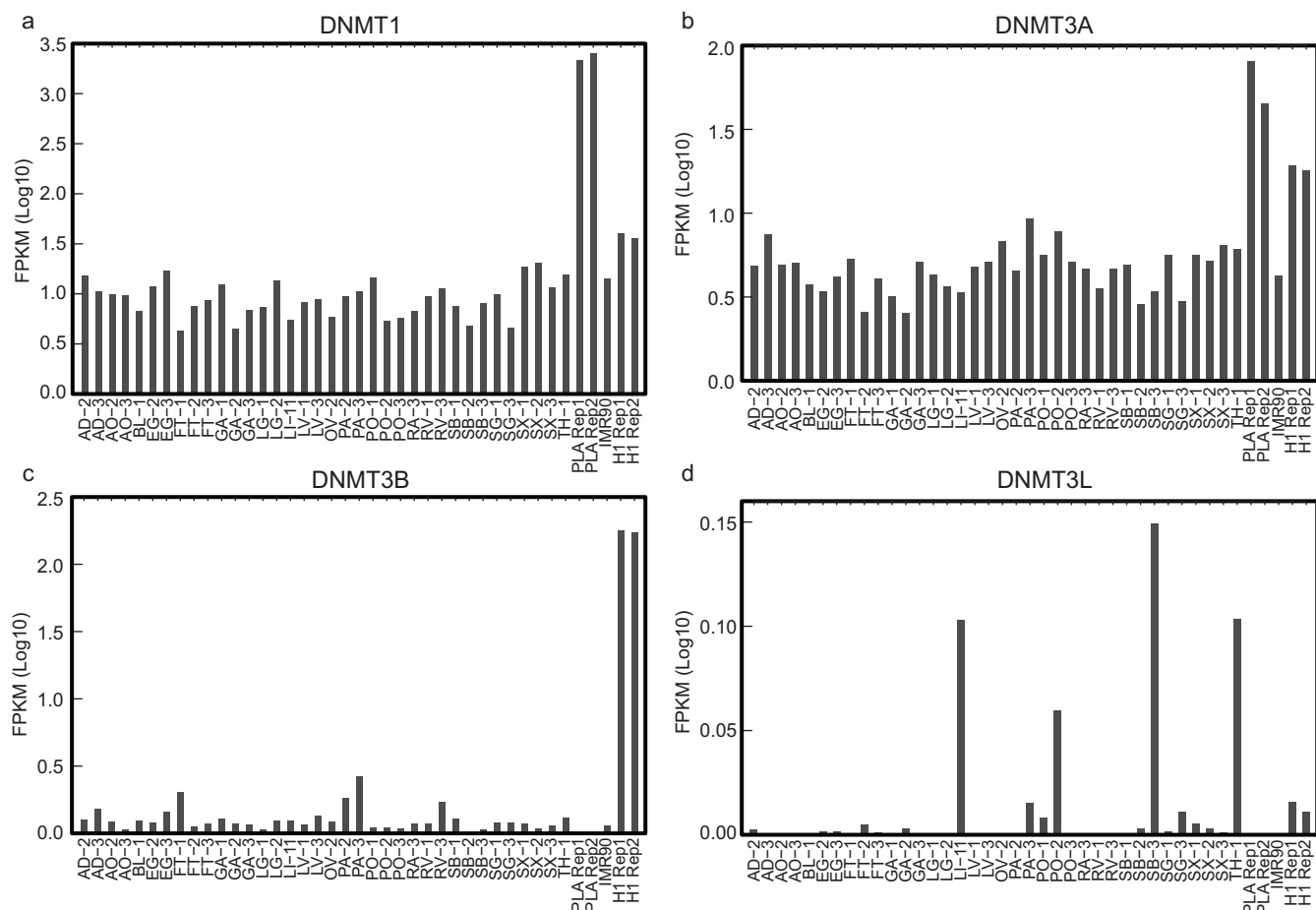
Extended Data Figure 7 | DNMT expression across tissues. **a–d**, Bar plots of the expression (measured in $\log_{10}$ FPKMs) of DNMT1 (**a**), DNMT3A (**b**), DNMT3B (**c**) and DNMT3L (**d**) across various samples.
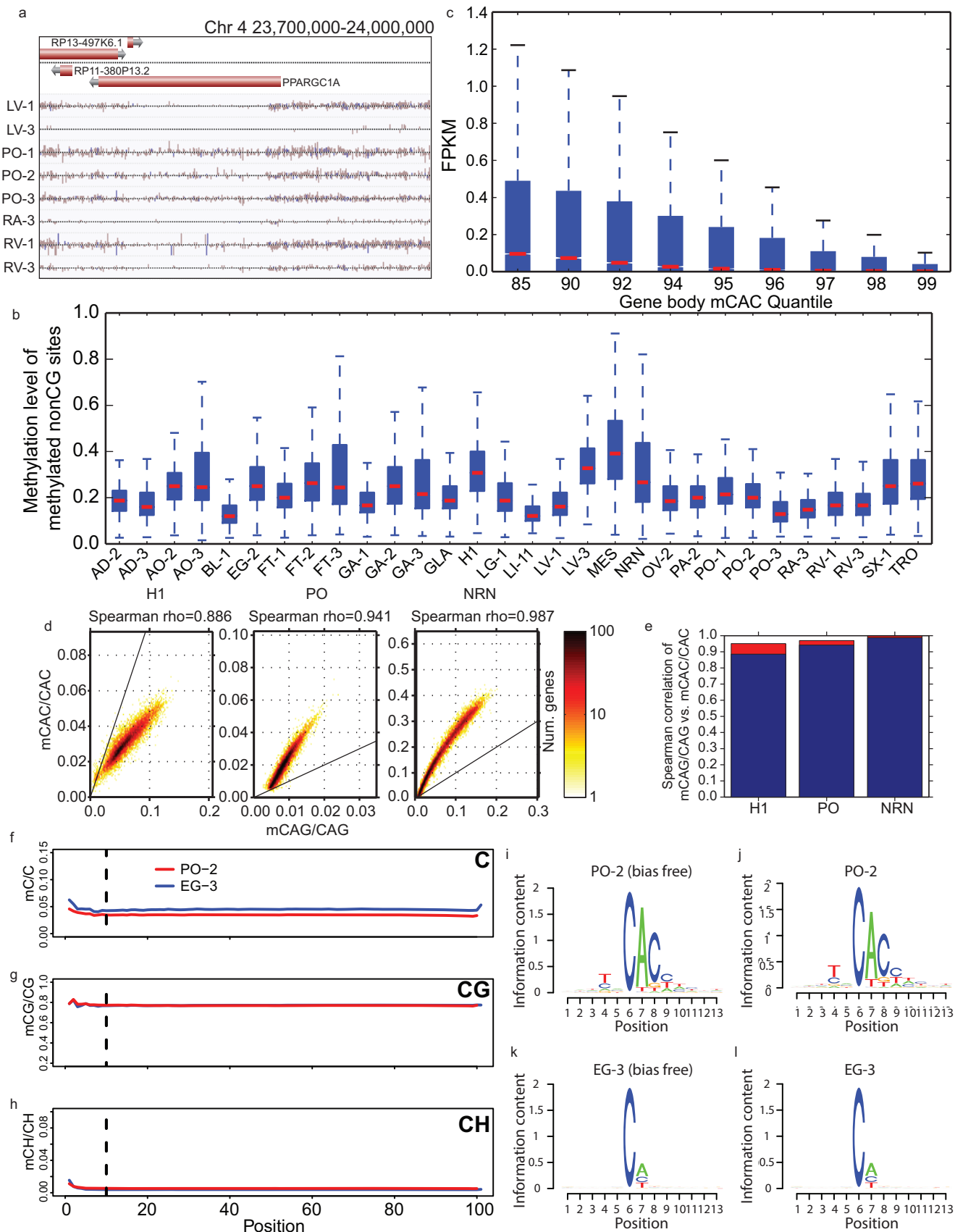
Extended Data Figure 8 | mCH distribution and correlation. a, A browser screenshot (see Fig. 1 for description) of an example region with non-CG methylation (mCH). Purple and pink ticks are methylated CHG and CHH sites, respectively (H = A, C or T). Ticks on the forward strand are projected upward from the dotted line and ticks on the reverse strand are projected downward. **b**, The distribution of methylation levels at mCH sites across all samples with a discernible TNCAC motif. Only mCH sites with at least 10 reads and a significant amount of methylation were considered. **c**, Boxplots of the expression values across different quantiles of CAC gene body methylation (gene body mCAC). **d**, Scatterplot of mCAG versus mCAC inside gene bodies. **e**, Bar plot of the correlation of mCAG and mCAC inside gene bodies (blue) and the theoretical maximal correlation (red) if mCAC and mCAG are perfectly correlated. **f–h**, The methylation levels of C (top), CG (middle) and CH (bottom) across the read positions for PO-2 (red line) and EG-3 (blue line). Vertical lines indicate the position (tenth base from the beginning) where trimming was applied. **i**, mCH motif from PO-2 with the first 10 bases of each read trimmed. **j**, mCH motif from PO-2 without trimming. **k**, mCH motif from EG-3 with the first 10 bases of each read trimmed. **l**, mCH motif from EG-3 without trimming. The height of each letter represents its information content (that is, prevalence).
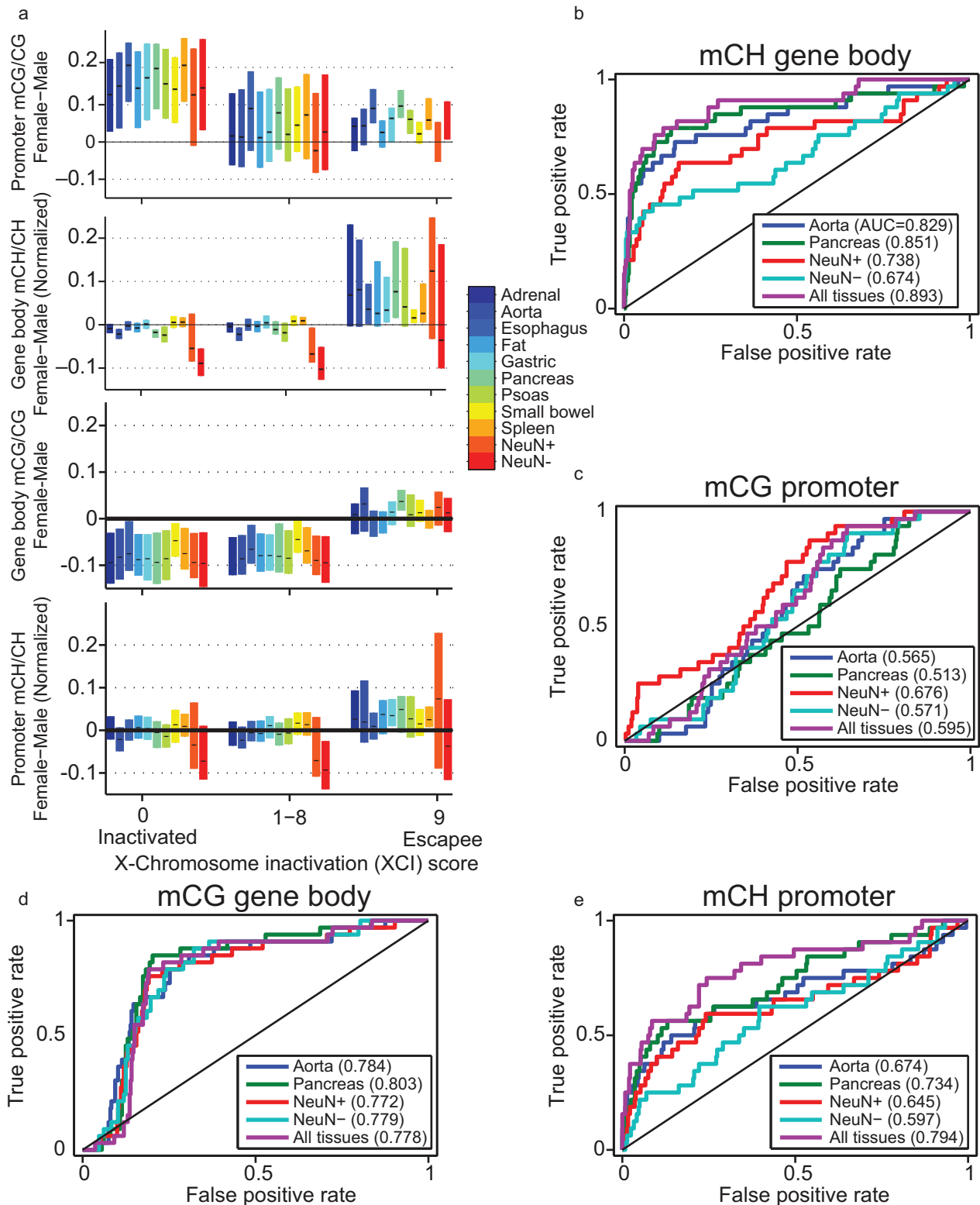
Extended Data Figure 9 | X-chromosome inactivation. a, Distributions of promoter CG methylation (mCG) levels (mCG/CG), gene body non-CG methylation (mCH) levels (mCH/CH), gene body mCG levels and promoter mCH levels in genes previously reported to express from only one allele (inactivated) or biallelically (escapee)[29]. Black ticks show median, and bars indicate the twenty-fifth to seventy-fifth percentile range. Genes more prone to escaping inactivation have lower promoter mCG, higher gene body mCH, higher gene body mCG and higher promoter mCH in females. **b–e**, Discriminability analysis using gender-specific gene body mCH (**b**), promoter mCG (**c**), gene body mCG (**d**) and promoter mCH (**e**) to predict the escapee status of X-linked genes, respectively. Among them, gene body mCH is the most predictive feature of X-chromosome inactivation escapees. The discriminability was measured by the area under the curve (AUC) (Supplementary Methods).
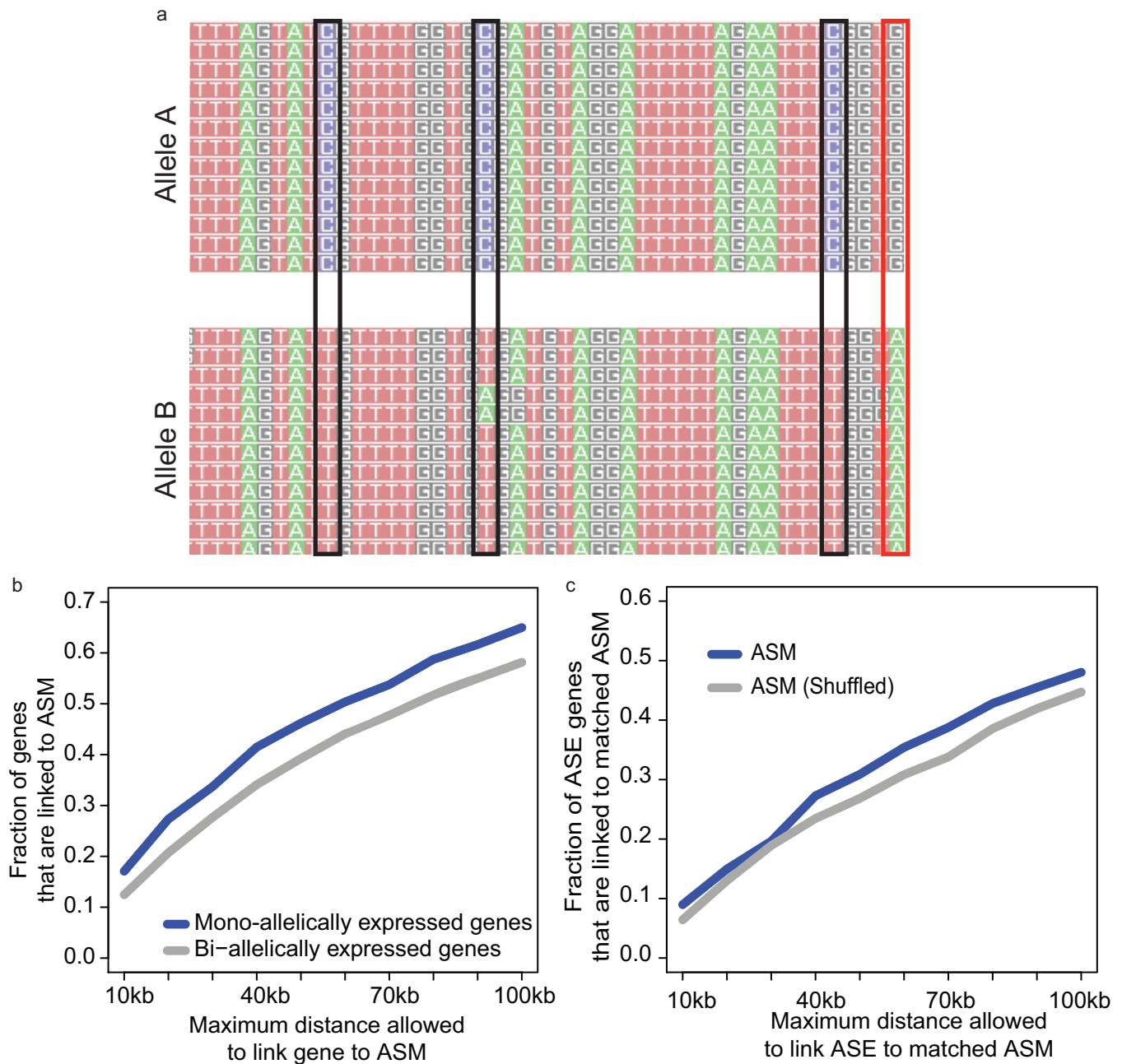
Extended Data Figure 10 | ASM and ASE. a, An example of ASM. Reads that contain a heterozygous SNP (red box) are separated by allele. The number of methylated (reads containing Cs) and unmethylated (reads containing Ts) at adjacent CG sites (black boxes) are tested for differential methylation. **b,** Fraction of ASE genes (blue) and bi-allelically expressed genes (grey) that have at least one ASM event within a certain distance. Bi-allelically expressed genes were defined as genes that were covered by at least 10 reads and whose *P* values, given by binomial test for allelic expression, were greater than 0.2 (that is, no significance). **c,** Fraction of ASE genes that were linked to matched ASM events (blue) and matched ASM events with their locations shuffled (grey). **b, c,** Aggregated results using samples from triplicate tissues.

LETTER

Global circulation patterns of seasonal influenza viruses vary with antigenic drift

Trevor Bedford[1], Steven Riley[2,3], Ian G. Barr[4], Shobha Broor[5], Mandeep Chadha[6], Nancy J. Cox[7], Rodney S. Daniels[8], C. Palani Gunasekaran[9], Aeron C. Hurt[4,10], Anne Kelso[4], Alexander Klimov[7]‡, Nicola S. Lewis[11], Xiyan Li[12], John W. McCauley[8], Takato Odagiri[13], Varsha Potdar[6], Andrew Rambaut[3,14,15], Yuelong Shu[12], Eugene Skepner[11], Derek J. Smith[11,16], Marc A. Suchard[17,18,19], Masato Tashiro[13], Dayan Wang[12], Xiyan Xu[7], Philippe Lemey[20] & Colin A. Russell[21]

Understanding the spatiotemporal patterns of emergence and circulation of new human seasonal influenza virus variants is a key scientific and public health challenge. The global circulation patterns of influenza A/H3N2 viruses are well characterized[1–7], but the patterns of A/H1N1 and B viruses have remained largely unexplored. Here we show that the global circulation patterns of A/H1N1 (up to 2009), B/Victoria, and B/Yamagata viruses differ substantially from those of A/H3N2 viruses, on the basis of analyses of 9,604 haemagglutinin sequences of human seasonal influenza viruses from 2000 to 2012. Whereas genetic variants of A/H3N2 viruses did not persist locally between epidemics and were reseeded from East and Southeast Asia, genetic variants of A/H1N1 and B viruses persisted across several seasons and exhibited complex global dynamics with East and Southeast Asia playing a limited role in disseminating new variants. The less frequent global movement of influenza A/H1N1 and B viruses coincided with slower rates of antigenic evolution, lower ages of infection, and smaller, less frequent epidemics compared to A/H3N2 viruses. Detailed epidemic models support differences in age of infection, combined with the less frequent travel of children, as probable drivers of the differences in the patterns of global circulation, suggesting a complex interaction between virus evolution, epidemiology, and human behaviour.

Owing to the frequency and severity of human seasonal influenza A/H3N2 virus epidemics, recent work has focused on the global circulation dynamics of H3N2 viruses[1–7]. Studies have shown that, each year, H3N2 epidemics worldwide result from the introduction of new genetic variants from East and Southeast (E-SE) Asia, where viruses circulate via a network of temporally overlapping epidemics[1,2,4,5], rather than local persistence[1,3,6,7]. In addition to H3N2, H1N1 viruses and two antigenically diverged lineages of influenza B viruses, B/Victoria/2/1987-like (Vic) and B/Yamagata/16/1988-like (Yam), circulate among humans with lower but substantial disease burdens[8,9]. Despite their importance, the global circulation dynamics of former seasonal H1N1 viruses (preceding the 2009 pandemic) and B viruses have been largely neglected.

Given that influenza A and B viruses cause similar symptoms and evolve by similar mechanisms of immune escape, we hypothesized that each would follow similar patterns of global circulation, with new genetic variants originating in East and Southeast Asia that rapidly replace existing genetic variants. To test this hypothesis we compared the global circulation patterns of the haemagglutinin (HA) genes of H3N2, former seasonal H1N1, Vic, and Yam viruses. We assembled data sets of HA sequences with complete HA1 domains for each subtype from the World Health Organization Global Influenza Surveillance and Response System and the Influenza Research Database[10], covering 2000–2012. To reduce the impact of surveillance biases, we subsampled these data to more equitable spatiotemporal distributions, resulting in data sets comprising 4,006 H3N2, 2,144 H1N1, 1,999 Vic, and 1,455 Yam HA sequences (Extended Data Fig. 1). Although deficient in viruses from Africa and Eastern Europe, to our knowledge these are the most geographically and temporally comprehensive seasonal influenza virus data sets assembled to date.

By estimating temporally resolved phylogenetic trees for each subtype, we revealed faster rates of nucleotide mutation and amino acid substitution in H3N2 and H1N1 than in the B viruses (consistent with previous work[11,12]), but more genealogical diversity in B viruses than in A viruses (Extended Data Table 1). This inverse relationship between evolutionary rate and genealogical diversity is expected if increased mutation rate correlates with antigenic drift[13] and drives increased adaptive evolution, thus purging HA genetic diversity[14]. By inferring geographic ancestry using Bayesian phylogeographic methods[15], we found a consistent pattern for H3N2 viruses (Fig. 1a) in which viruses worldwide rapidly coalesce to the trunk of the tree (average time to trunk = 1.42 years), with trunk viruses mostly originating from East and Southeast Asia (Extended Data Fig. 2a). This finding is consistent with previously reported patterns[1,2,4,5], with East and Southeast Asia acting as the source population for epidemics worldwide.

In addition to China and Southeast Asia, India frequently contributed viruses to the trunk of the tree, suggesting that the global circulation of H3N2 viruses is maintained by an East and Southeast Asian network that includes India. India's role in the global dissemination of H3N2 viruses may have been similar historically, but India-wide influenza surveillance only began in 2004. There were brief periods, notably the 2007–2008 Northern Hemisphere winter, when regions outside

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [2]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London SW7 2AZ, UK. [3]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. [4]World Health Organization (WHO) Collaborating Centre for Reference and Research on Influenza, Melbourne, Victoria 3000, Australia. [5]SGT Medical College, Hospital and Research Institute, Village Budhera, District Gurgaon, Haryana 122505, India. [6]National Institute of Virology, Pune 411001, India. [7]WHO Collaborating Center for Reference and Research on Influenza, Centers for Disease Control and Prevention, Atlanta, Georgia 30329, USA. [8]WHO Collaborating Center for Reference and Research on Influenza, Medical Research Council National Institute for Medical Research (NIMR), London NW7 1AA, UK. [9]King Institute of Preventive Medicine and Research, Guindy, Chennai 600032, India. [10]Melbourne School of Population and Global Health, University of Melbourne, Parkville, Victoria 3010, Australia. [11]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [12]WHO Collaborating Center for Reference and Research on Influenza, China CDC, Beijing 102206, China. [13]WHO Collaborating Center for Reference and Research on Influenza, National Institute of Infectious Diseases, Tokyo 208-0011, Japan. [14]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [15]Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3FL, UK. [16]Department of Viroscience, Erasmus Medical Center, 3015 Rotterdam, The Netherlands. [17]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, California 90095, USA. [18]Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California 90095, USA. [19]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California 90095, USA. [20]Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, 3000 Leuven, Belgium. [21]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK.
‡Deceased.

©2015 Macmillan Publishers Limited. All rights reserved
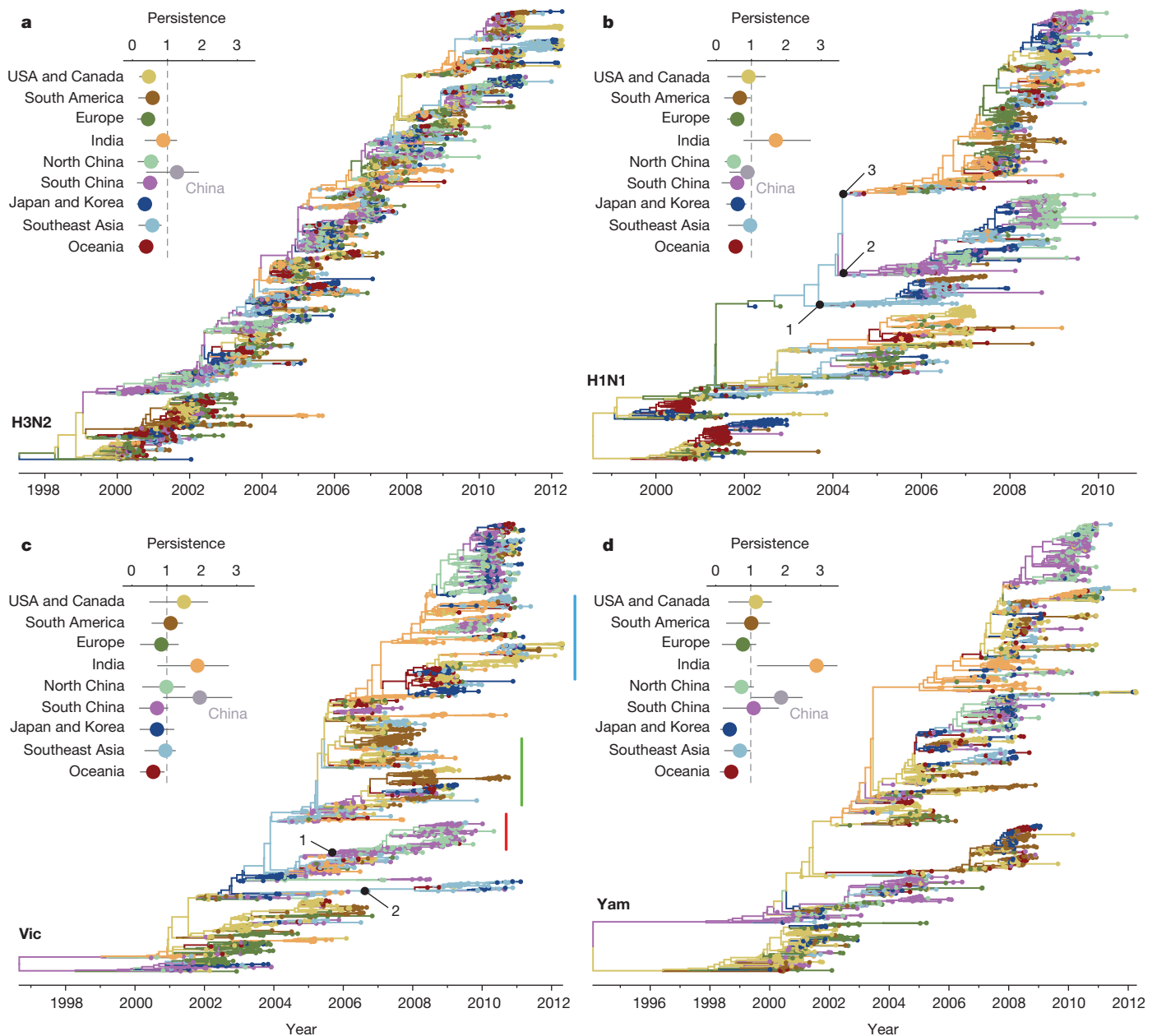
Figure 1 | Maximum clade credibility trees. a–d, Trees created with primary data sets of 4,006 H3N2 viruses (**a**), 2,144 H1N1 viruses (**b**), 1,999 Vic viruses (**c**) and 1,455 Yam viruses (**d**). Branch tips are coloured by geographic region of virus collection; internal branches are coloured by geographic region as inferred by Bayesian phylogeographic methods (region colours in persistence insets). In **b**, nodes 1–3 indicate co-circulating clades that diverged in 2004. In **c**, nodes 1 and 2 indicate divergent clades of viruses from Asia, coloured vertical bars indicate antigenic variants shown in Extended Data

Fig. 5a (green, B/Malaysia/2506/2004-like; red, B/Hubei Songzi/52/2008-like; other post-2008 viruses, B/Brisbane/60/2008-like). The inset to the top left of each tree shows duration of region-specific persistence measured as the waiting time in years for a virus to leave its geographic region of origin. Circles represent mean persistence across sampled viruses, while lines show the interquartile range of persistence across sampled viruses. Region 'China' shows the combined persistence estimate for North China and South China together.

East and Southeast Asia contributed to the trunk of the H3N2 tree. However, these instances were rare and trunk viruses from outside East and Southeast Asia descended directly from viruses within East and Southeast Asia (Fig. 1a). Quantifying the average ancestry of strains from each geographic region in the 3 years before sampling showed prominent roles for China, India, and Southeast Asia in seeding epidemics in all regions (Extended Data Fig. 3).

Surprisingly, the global circulation patterns of former seasonal H1N1 viruses differed substantially from those observed for H3N2 viruses (Fig. 1). Like H3N2, most lineages of H1N1 viruses eventually coalesced with viruses from East and Southeast Asia and India. However, this coalescence was slower than for H3N2 viruses with

prolonged co-circulation of geographically segregated H1N1 lineages (Fig. 1b, Extended Data Figs 3 and 4). Geographic segregation of H1N1 viruses was particularly pronounced beginning in 2004/2005, with the emergence of three co-circulating genetic lineages (Fig. 1b, nodes 1–3) that each independently acquired HA mutations leading to antigenic evolution from the A/New Caledonia/20/1999-like phenotype to the A/Solomon Islands/3/2006-like phenotype. These lineages circulated in Southeast Asia (node 1), China (node 2) and India (node 3), with the Indian lineage eventually spreading worldwide before the emergence of H1N1pdm09 viruses.

Phylogeographic analyses of B Vic and Yam viruses revealed further differences from H3N2 viruses with lineages frequently circulating

outside of East and Southeast Asia for several years without evidence of seeding from East and Southeast Asia (Fig. 1c, d). Prominent examples include the seeding of the North American 2006/2007 Vic season directly from 2005/2006 North American viruses and the seeding of the North American 2001/2002 Yam season directly from 2000/2001 North American viruses (Extended Data Fig. 4). Similarly, lineages of viruses within East and Southeast Asia commonly circulated exclusively in East and Southeast Asia for more than 1 year. These long circulating East and Southeast Asian lineages were most apparent for Vic viruses where two lineages (Fig. 1c, nodes 1 and 2) persisted independently in China and Southeast Asia for over 5 years without spreading to other regions and led to the co-circulation of three distinct Vic antigenic variants in different parts of the world during 2007–2008 (Extended Data Fig. 5a).

Patterns of persistence of genetic variants differed by subtype and region, with H3N2 viruses persisting regionally for an average of ~6 months, H1N1 for ~9 months, Vic for ~13 months and Yam for ~12 months. H3N2 viruses showed comparably short durations of persistence across the world (Fig. 1), with the exceptions of India and China. Patterns within China were characterized by North and South lineages contributing jointly to persistence, as combining North and South phylogeny nodes resulted in substantially greater persistence estimates than from North or South lineages alone (Fig. 1). For H3N2, evidence for joint contributions to persistence by region pairs that exclude China is comparatively weak (Extended Data Fig. 6a, Supplementary Information). For Vic and Yam, the mean duration of persistence was longer than for H3N2 or H1N1 in most regions, particularly in India and China where mean durations were >2 years (Fig. 1, Extended Data Fig. 4). Duration of regional persistence correlated with the proportion of virus originating from that region (Extended Data Fig. 6b) and observed phylogeographic patterns were robust to subsampling assumptions (Supplementary Information, Extended Data Table 2).

To investigate differences in the global migration patterns of H3N2, H1N1 and B viruses, we used the spatiotemporally resolved phylogenies to estimate the amounts of virus movement between regions (Fig. 2). Rates of movement between pairs of regions were highly correlated between viruses with Spearman correlation coefficients ranging from 0.65 (H3N2 vs Yam) to 0.75 (H3N2 vs H1N1), suggesting similar global connectivity networks for all viruses. However, while the overall structure of the migration network was similar, H3N2 viruses moved between regions more frequently than H1N1 and B viruses (migration events per lineage per year H3N2 = 1.96, H1N1 = 1.27, Vic = 0.93, Yam = 0.97, Extended Data Table 1).

We hypothesized a relationship between rates of global movement and rates of antigenic drift: although rates of genetic evolution were similar for H3N2 and H1N1 viruses, both H1N1 and B viruses evolved antigenically more slowly than H3N2 viruses[13] (Extended Data

Table 1). We also hypothesized that lower rates of immune escape for B and H1N1 compared with H3N2 would lead to younger average ages of infection, as children increasingly comprise the largest pool of susceptible individuals, and smaller, less frequent epidemics owing to smaller populations of susceptible individuals[13]. These differences are consistent with results from several community-based cohort studies that found that children were more frequently infected with B viruses than adults[8,16,17]. Age of infection data covering 2002–2011 from Australia show that H1N1 and B viruses infect younger individuals than H3N2 viruses (Extended Data Fig. 5b–d, median age of infection H3N2 = 30 years, H1N1 = 20 years, B = 16 years) and epidemiological data from Australia and the United States show reduced size and frequency of H1N1 and B epidemics compared to H3N2 (Extended Data Fig. 5f–i).

Differences in age of infection may explain differences in global circulation as children travel long distances much less frequently than adults (Extended Data Fig. 5e). A previous study hypothesized that age-specific patterns of infection could lead to differences in contact rates and the spread of influenza types within the United States over the course of a single season[18]. Here, we hypothesized that differential global air travel provides a plausible mechanism by which H1N1 and B

Figure 2 | Estimates of mean pairwise virus migration rate. Line thickness between regions indicates average number of migration events per lineage per year. Arrowhead size indicates the strength of directionality of migration. For clarity, only arrows corresponding to migration rates greater than 0.25 events per lineage per year are shown. Circle area indicates the global proportion of ancestry deriving from each region.

Figure 3 | Relationship of antigenic drift to incidence (a), **proportion of childhood infections** (b), **and geographic migration rate** (c), **in a multi-strain multi-region model of influenza transmission.** Black points represent outcomes from a model in which children and adults travel between regions at equal rates. Red points represent outcomes from a model in which adults travel between regions at 5.26× the rate of children (Extended Data Fig. 5e). Solid black and red lines represent LOESS fits to the data. With 2 travel scenarios, 7 mutation rates and 8 replicates, there are 112 individual stochastic simulations (Extended Data Fig. 7). Antigenic drift was measured in cartographic units[13] per year (see Methods). In a, attack rate was measured as proportion of the total population infected yearly. In c, migration rate was measured in terms of migration events per lineage per year.

viruses show increased genetic differentiation and reduced rates of global migration across multiple seasons, compared to H3N2 viruses.

To test the impact of differences in age distribution of infection on global patterns of virus movement, we constructed a multi-patch transmission model. We modelled two scenarios for host movement: (1) age-independent mixing between patches; (2) age-stratified mixing with host movement derived from air travel passenger age data (Extended Data Fig. 5e). In the age-independent scenario, model parameters only differed in rate of antigenic mutation, leading to differences in observed rates of antigenic drift among viruses and hence epidemic size and frequency (Extended Data Fig. 7). Faster antigenic drift resulted in greater incidence and more adult infections (Fig. 3a, b), but only modest differences in virus lineage movement (Fig. 3c, B-like viruses differ from H3-like viruses by a factor of 1.2), consistent with slightly faster spread of antigenically novel strains. However, age-stratified mixing between patches intensified the effect of antigenic drift on migration rate and created differences in rates of movement between patches more consistent with those observed for H3N2 vs H1N1 and B (Fig. 3c, B-like viruses differ from H3-like by a factor of 1.6). In the scenario with faster antigenic drift, infections were more mobile owing to greater frequency of adult infection, causing a knock-on effect on rates of viral movement. The model also suggests that the differences in patterns of regional persistence observed in the phylogenies might be shaped by a combination of differences in rates of antigenic evolution and variation in amplitude of epidemic seasonality, with slowly evolving viruses persisting longer than rapidly evolving viruses at low amplitudes of seasonal forcing (Extended Data Fig. 8a, Supplementary Information).

In the model, varying transmission rate rather than antigenic mutation rate also resulted in differences in the observed rate of antigenic drift, with higher transmission resulting in faster drift (Extended Data Fig. 8b). The relationship between antigenic drift rate and migration rate is similar, regardless of whether drift is modulated by mutation rate or transmission rate (Extended Data Fig. 8b). This finding is in line with theoretical work showing that epidemiological processes can influence influenza virus evolution[19,20]. However, there are important virological differences between influenza viruses that are likely to affect the efficiency and tempo at which antigenic variation is generated and fixed, which could in turn affect epidemiology[21–24] (Supplementary Information).

Regardless of the underlying drivers, there is a remarkable correspondence in model behaviour, quantified as a stable relationship between observable rate of antigenic drift and global circulation patterns. The patterns of epidemic spread observed here suggest that differences in ages of infection could explain patterns of global circulation across a variety of human viruses.

1. Russell, C. A. et al. The global circulation of seasonal influenza A (H3N2) viruses. Science **320,** 340–346 (2008).
2. Lemey, P. et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog. **10,** e1003932 (2014).
3. Rambaut, A. et al. The genomic and epidemiological dynamics of human influenza A virus. Nature **453,** 615–619 (2008).
4. Bedford, T., Cobey, S., Beerli, P. & Pascual, M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). PLoS Pathog. **6,** e1000918 (2010).
5. Chan, J., Holmes, A. & Rabadan, R. Network analysis of global influenza spread. PLOS Comput. Biol. **6,** e1001005 (2010).
6. Nelson, M. I., Simonsen, L., Viboud, C., Miller, M. A. & Holmes, E. C. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. PLoS Pathog. **3,** e131 (2007).
7. Nelson, M. I. et al. Stochastic processes are key determinants of short-term evolution in influenza a virus. PLoS Pathog. **2,** e125 (2006).
8. Glezen, P. W., Schmier, J. K., Kuehn, C. M., Ryan, K. J. & Oxford, J. The burden of influenza B: a structured literature review. Am. J. Public Health **103,** e43–e51 (2013).
9. Thompson, W. W. et al. Influenza-associated hospitalizations in the United States. J. Am. Med. Assoc. **292,** 1333–1340 (2004).
10. Squires, R. B. et al. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. Influenza Other Respir. Viruses **6,** 404–416 (2012).
11. Chen, R. & Holmes, E. C. The evolutionary dynamics of human influenza B virus. J. Mol. Evol. **66,** 655–663 (2008).
12. Cox, N. J. & Bender, C. A. The molecular epidemiology of influenza viruses. Semin. Virol. **6,** 359–370 (1995).
13. Bedford, T. et al. Integrating influenza antigenic dynamics with molecular evolution. eLife **3,** e01914 (2014).
14. Bedford, T., Cobey, S. & Pascual, M. Strength and tempo of selection revealed in viral gene genealogies. BMC Evol. Biol. **11,** 220 (2011).
15. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. PLOS Comput. Biol. **5,** e1000520 (2009).
16. Fox, J. P., Hall, C. E., Cooney, M. K. & Foy, H. M. Influenzavirus infections in Seattle families, 1975–1979. I. Study design, methods and the occurrence of infections by time and age. Am. J. Epidemiol. **116,** 212–227 (1982).
17. Longini, I. M. Jr, Koopman, J. S., Monto, A. S. & Fox, J. P. Estimating household and community transmission parameters for influenza. Am. J. Epidemiol. **115,** 736–751 (1982).
18. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science **312,** 447–451 (2006).
19. Recker, M., Pybus, O. G., Nee, S. & Gupta, S. The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. Proc. Natl Acad. Sci. USA **104,** 7711–7716 (2007).
20. Zinder, D., Bedford, T., Gupta, S. & Pascual, M. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. PLoS Pathog. **9,** e1003104 (2013).
21. Nobusawa, E. & Sato, K. Comparison of the mutation rates of human influenza A and B viruses. J. Virol. **80,** 3675–3678 (2006).
22. Neuzil, K. M. et al. Immunogenicity and reactogenicity of 1 versus 2 doses of trivalent inactivated influenza vaccine in vaccine-naive 5–8-year-old children. J. Infect. Dis. **194,** 1032–1039 (2006).
23. Matrosovich, M. N. et al. Probing of the receptor-binding sites of the H1 and H3 influenza A and influenza B virus hemagglutinins by synthetic and natural sialosides. Virology **196,** 111–121 (1993).
24. Hensley, S. et al. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science **326,** 734–736 (2009).

Author Contributions C.A.R. and T.B. conceived the research. C.A.R. and T.B. drafted the manuscript with substantial support from P.L. and S.R. I.G.B., S.B., M.C., N.J.C., R.S.D., C.P.G., A.C.H., A.K., A.Kl. X.L., J.W.M., T.O., V.P., Y.S., M.T., D.W. and X.X. coordinated and produced the influenza surveillance data. T.B. performed the modeling and data analyses along with C.A.R., S.R., P.L., M.A.S. and A.R. T.B. created the figures. All authors discussed the results and contributed to the revision of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.R. (car44@cam.ac.uk).

METHODS

Sequence data. Haemagglutinin (HA) coding sequences for influenza A H3N2 viruses, former seasonal H1N1 viruses (preceding the 2009 pandemic), and influenza B virus lineages Victoria (Vic) and Yamagata (Yam) collected by the World Health Organization (WHO) Global Influenza Surveillance and Response Network including the National Institute of Virology, Pune, India between 2000 and 2012 were combined with human seasonal influenza virus sequences (minimum length = 984 base pairs) covering 2000 to 2012 from the Influenza Research Database[10]. After removing duplicate strains and strains overly divergent based on root-to-tip distances, the data set contained 9,139 H3N2 sequences, 3,789 H1N1 sequences, 2,577 Vic sequences and 1,821 Yam sequences. Sampling locations for these sequences were parsed from strain names. Sequences were grouped into 9 geographic regions: USA/Canada, South America, Europe, India, North China, South China, Japan/Korea, Southeast Asia and Oceania. Specifics of this partitioning are shown in Extended Data Fig. 1. Groups were chosen to maximize available sequences within each region while still providing enough geographic diversity to ensure nearly global coverage. Sequences from Africa, Central America, the Middle East and Russia were excluded because of a lack of sufficient numbers of sequences to provide comparable estimates to other regions.

In the raw sequence data, some regions, such as the USA, were over-represented. Additionally, more recent years were over-represented compared to years at the start of the study period. In order to control for these sampling biases, we subsampled the raw data randomly by location and time to create a more equitable spatiotemporal distribution. The USA had consistently more sequences available every year from 2000 to 2012, thus in order to maintain similar total numbers of sequences for each region across the entire study period it was necessary to sample fewer sequences per year from the USA. We selected 50 sequences per region per year (40 for USA/Canada) for H3N2 and 80 sequences per region per year (45 for USA/Canada) for H1N1, Vic and Yam. This subsampling resulted in largely similar sequence counts across years and across regions for each virus, but overall more H3N2 sequences than H1N1 or B sequences, with 4,006 H3N2 sequences, 2,144 H1N1 sequences, 1,999 Vic sequences and 1,455 Yam sequences (Extended Data Fig. 1). When selecting subsampled sequences we first selected sequences with full day-month-year collection dates and then longer sequences over sequences with less precise dates or shorter sequences. HA sequence data for 1,630 H3N2 isolates, 1,600 H1N1 isolates, 1,394 Vic isolates and 881 Yam isolates have been deposited in the Influenza Research Database[10] and accession numbers for all sequences used provided as Supplementary Information.

Phylogeographic inference. Time-resolved phylogenetic trees were estimated for H3N2, H1N1, Vic and Yam using BEAST v1.8.1[25] and incorporated the SRD06 nucleotide substitution model[26], a coalescent demographic model with constant effective population size and a strict molecular clock across branches. A strict molecular clock was chosen based on finding strong correlations between date of sampling and evolutionary distance in all data sets, as estimated by Path-O-Gen v1.4 (http://tree.bio.ed.ac.uk/software/pathogen/). Using a strict clock also reduced the risk of model over-parameterization (for example, for the complete H3N2 data set with a relaxed clock, there would be $2 \times 4{,}006 - 2 = 8{,}010$ branch-specific rates). Samples with imprecise dates (known only to the month or to the year) had their dates of sampling estimated assuming a uniform prior within the known temporal bounds[27]. Markov Chain Monte Carlo (MCMC) was run for 600 million steps and trees were sampled every 5 million steps after allowing a burn-in of 100 million steps, yielding a total sample of 100 trees for H1N1, Vic and Yam. With significantly more samples, H3N2 required a longer chain to converge. Here, MCMC was run in parallel for 2 chains, each with 650 million steps sampled every 3 million steps with a burn-in of 500 million steps and samples across chains combined, yielding a total of 100 sampled trees. These trees were treated as independent draws from the posterior space of trees when subsequently used in the robust counting and phylogeographic analyses[28]. Evolutionary rates in Extended Data Table 1 were estimated using the 'renaissance' counting methods of Lemey et al.[29].

Phylogeographic patterns were estimated using a discrete-state continuous time Markov chain (CTMC) model, in which transition rates were estimated between each pair of regions[15]. We assumed a non-reversible transition model[30] consisting of 72 separate rate parameters, each with a Bayesian stochastic search variable selection (BSSVS) indicator variable, and a separate overall rate of geographic transition. We assumed an exponential prior with mean of 1 for each transition rate, a negative binomial prior with mean of 9 and standard deviation of 9 for the total number of non-zero rates and an exponential prior with mean of 1 migration event per lineage per year for the overall geographic transition rate. MCMC was run for 12 million steps with a burn-in of 2 million steps, and parameters sampled every 10,000 steps and trees sampled every 100,000 steps, yielding a total sample of 1,000 parameter states and 100 trees on which estimates were based. Pairwise migration rate estimates had an effective sample size (ESS) of 350 at the minimum and most had ESS greater than 500.

This procedure yielded posterior trees with the geographic states of internal nodes resolved. We analysed these posterior trees using the program PACT v0.9.5 (https://github.com/trvrb/PACT) to compute the following summary statistics: (1) genealogical diversity[14], measuring the average time it takes for two randomly chosen contemporaneous lineages to coalesce, (2) time to the most recent common ancestor (TMRCA)[14], measuring the average time it takes for all contemporaneous lineages to find a common ancestor, (3) genealogical F_{ST}, measuring the degree of population structure in contemporaneous lineages calculated as $F_{ST} = (\pi_b - \pi_w)/\pi_b$, where π_w is genealogical diversity between randomly sampled lineages from the same geographic region and π_b is genealogical diversity between randomly sampled lineages from different geographic regions, (4) persistence, measuring the average number of years for a tip to leave its sampled location, walking backwards up the phylogeny, (5) migration rate, measuring the average number of migration events over the phylogeny divided by total tree length to give migration events per lineage per year, (6) trunk location through time[4], measuring the posterior distribution across sampled phylogenies of the trunk geographic state, where the trunk is defined as all branches ancestral to viruses sampled within 1 year of the most recent sample, (7) region-specific ancestral geographic history, measuring the distribution of geographic locations of tips belonging to a particular region traced backwards in time through the phylogeny averaged across sampled phylogenies. Statistics (1), (2), (3), (6), and (7) were calculated across 0.1 year genealogical windows. These procedures gave an estimate of credible intervals for inferred ancestral locations across posterior phylogeographic reconstructions.

Code and data availability. Sequence data has been deposited with the Influenza Research Database[10] and accession numbers provided as Supplementary Data. The entire bioinformatic pipeline, including data subsampling, preparing XML files for BEAST, setting up PACT analyses and rendering figures is available at https://github.com/blab/global-migration. Analysis and data files are archived on the Dryad Digital Repository under DOI http://dx.doi.org/10.5061/dryad.pc641.

Surveillance, travel and age-structure data. We investigated epidemic size and frequency using virological isolation data between 2000 and 2012 collected by the WHO Collaborating Centre for Reference and Research on Influenza at the Victorian Infectious Diseases Reference Laboratory (VIDRL), Melbourne, Australia and the Centers for Disease Control and Prevention, Atlanta, USA (Extended Data Fig. 5f–i). These isolations were categorized by date of sampling and by virus type: H3N2, H1N1, Vic, or Yam. The data from VIDRL also contained information on patient age. The age structure of incidence was estimated by constructing a distribution of age of infection from individuals > 5 years (owing to the overrepresentation of < 5 year old patients for all subtypes) (Extended Data Fig. 5b–d). Median age of infection was 30 years (H3N2), 20 years (H1N1) and 16 years (B) and mean age of infection was 33.9 years (H3N2), 23.1 years (H1N1) and 23.2 years (B). Median age of infection was significantly different for H3N2 vs H1N1 ($P = 4.6 \times 10^{-29}$, Mann–Whitney U test), H3N2 vs B ($P = 1.2 \times 10^{-62}$) and H1N1 vs B ($P = 0.041$). The patient age data from VIDRL were potentially biased by testing strategy and the generally higher severity of H3N2 virus infections. Children and working age adults were more likely to be tested than the elderly but the greater severity of H3N2 virus infections might spread and flatten the patient age distribution. For this reason we additionally tested excluding individuals > 65 years and recalculating summary statistics, finding median ages of infection of 27 years (H3N2), 19 years (H1N1) and 15 years (B) and mean age of infection as 28.0 years (H3N2), 22.2 years (H1N1) and 20.3 years (B). We classified children as 0–15 years and adults as 16 years and older, and estimated proportion of childhood infections as 30% (H3N2), 52% (H1N1) and 60% (B). There are potentially other biases specific to individual sentinel physicians and hospitals that could affect sample collection. However, the estimate derived from the VIDRL data that ~60% of influenza B virus infections occur in children is consistent with other estimates (reviewed in Glezen et al.[8]). Other studies similarly corroborate the estimates of lower age of infection for H1N1 viruses as compared to H3N2[31,32].

Additionally, we analysed the distribution of ages of ~102.5 million air passengers travelling through London Heathrow and London Gatwick airports in 2011 (Extended Data Fig. 5E) reported by Civil Aviation Authority of the UK (http://www.caa.co.uk/docs/81/2011CAAPaxSurveyReport.pdf). Assuming that children of ages 0 to 15 make up 17% of the UK population (Office of National Statistics), this distribution suggests that children engage in air travel at 19% the rate of adults.

For the modelling described below, we estimated age-structured contact rates following the empirical mixing data provided by Mossong et al.[33]. These contact matrices were previously validated in modelling pertussis epidemiology[34]. We simplified the Mossong et al. mixing matrices to record child-to-child contacts, child-to-adult contacts, adult-to-child contacts and adult-to-adult contacts, where

children were defined to be 0 to 15 and adults to be 16 or over. This resulted in the following mixing matrix

$$\alpha = \begin{pmatrix} 1.0 & 0.21 \\ 0.21 & 0.26 \end{pmatrix},$$

where rates are relative to child-to-child contact rates.

Epidemiological modelling. An individual-based model of influenza evolution and epidemiology was constructed following methods presented in Bedford *et al.*[35]. The model used here is identical to Bedford *et al.* except where specified below. The present implementation used a linear-strain space[36,37], in which virus phenotype is represented by a continuous variable and cross-immunity between viruses is a function of distance between viruses in this space. We parameterized the model to compare scenarios of age-structured mixing between regions and to compare viruses with different rates of antigenic drift.

The model was simulated for 120 years with daily time steps and the first 100 years discarded to allow equilibrium to be reached. We modelled a metapopulation with individuals equally divided into three regions (North, Tropics, South). Individual's ages were tracked throughout the simulation and those less than 16 years old were classified as children and those 16 or older were classified as adults. Transmission occurred by mass action, with transmission rates modified by regional compartment and by age compartment. Thus, for example, the force of infection into children in the Tropics followed

$$\lambda_{ct} = \sum_{i \in (a,c)} \beta_t \alpha_{ic} I_{it} \frac{S_{ct}}{N_t} + \sum_{i \in (a,c)} \sum_{j \in (n,s)} \beta_t \alpha_{ic} m_i I_{ij} \frac{S_{ct}}{N_t},$$

where β_j is the seasonally forced contact rate in region j, α_{ac} represents adult-to-child transmission, m_i represents between-region transmission in age class i, I_{ij} represents the number of persons infected in age class i in region j, S_{ij} represents the number of susceptible persons in age class i in region j, and N_j represents the total number of hosts in region j. The northern and southern regions were seasonally forced in opposite phase with a sinusoidal function following ε, while the tropics had no seasonal forcing.

Each virus possessed a one-dimensional antigenic phenotype ϕ_v and after recovery a host 'remembered' its infecting phenotype. For each contact event, the Euclidean distance from infecting phenotype ϕ_v was calculated to each of the phenotypes in the host immune history $\phi_{h_1},\ldots,\phi_{h_n}$. Here, one unit of antigenic distance was designed to roughly correspond to a twofold dilution of antiserum in a haemagglutination inhibition (HI) assay[38]. The probability ρ that infection occurred after exposure was proportional to the distance d to the closest phenotype in the host immune history, following $\rho = \min\{d\,s, 1\}$. Each day there was a chance μ that an infection mutates to a new phenotype. This mutation rate represents a phenotypic rate, rather than genetic mutation rate, and can be thought of as arising from multiple genetic sources. When a mutation occurred, the virus's phenotype was moved either left or right randomly and mutation size sampled from an exponential distribution with mean step size σ_{avg}. Epidemiological parameters for the baseline epidemiological scenario with notation following Bedford *et al.*[35] were:

- Base transmission rate $\beta = 0.88$ per day
- Duration of infection $1/v = 5$ days
- Birth/death rate $= 1/50$ years
- Total population size $N = 45$ million
- Seasonal forcing in north and south $\varepsilon = 0.15$
- Antigenic scaling $s = 0.07$
- Antigenic mutation rate $\mu = 0.5$ to 6.5×10^{-4} per day

- Average mutation size $\sigma_{avg} = 0.3$ units
- Child-to-child transmission $\alpha_{cc} = 1.00$
- Child-to-adult transmission $\alpha_{ca} = 0.21$
- Adult-to-child transmission $\alpha_{ac} = 0.21$
- Adult-to-adult transmission $\alpha_{aa} = 0.26$
- Child between-region transmission $m_c = 0.0020$
- Adult between-region transmission $m_a = 0.0020$

In the model with age-stratified mixing with host movement derived from air travel passenger age data, child between-region transmission m_c was 0.0011 and adult between-region transmission m_a was 0.0060.

In the course of the simulation, the underlying infection history of who infects whom was recorded and output as a complete infection tree. Without ample within-host diversity owing to chronic infection, the complete infection tree also generated a fully observed phylogenetic tree. Examining geographic location across the phylogenetic tree allowed us to directly calculate migration rate as total migration events observed (transitions from one region to another) divided by total opportunity (tree length).

The simulation was parameterized to model H3-like, H1-like and B-like behaviour (Extended Data Fig. 7) by modulating antigenic mutation rate μ in the primary analysis (Fig. 3) or transmission rate β as a secondary analysis (Extended Data Fig. 8b). Values for μ and β were chosen based on observed attack rate, proportion of childhood infections, and antigenic drift rate.

Source code for the simulation is available at https://github.com/trvrb/antigen/tree/global-migration and parameter and results files are available at https://github.com/blab/global-migration/tree/master/model.

25. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
26. Shapiro, B., Rambaut, A. & Drummond, A. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23,** 7–9 (2006).
27. Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* **28,** 879–887 (2011).
28. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53,** 673–684 (2004).
29. Lemey, P., Minin, V. N., Bielejec, F., Pond, S. L. K. & Suchard, M. A. A counting renaissance: combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. *Bioinformatics* **28,** 3248–3256 (2012).
30. Edwards, C. J. *et al.* Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol.* **21,** 1251–1258 (2011).
31. Kelly, H. A., Grant, K., Williams, S., Fielding, J. & Smith, D. Epidemiological characteristics of pandemic influenza H1N1 2009 and seasonal influenza infection. *Med. J. Aust.* **191,** 146–149 (2009).
32. Khiabanian, H., Farrell, G., St George, K. & Rabadan, R. Differences in patient age distribution between influenza A subtypes. *PLoS ONE* **4,** e6832 (2009).
33. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5,** e74 (2008).
34. Rohani, P., Zhong, X. & King, A. A. Contact network structure explains the changing epidemiology of pertussis. *Science* **330,** 982–985 (2010).
35. Bedford, T., Rambaut, A. & Pascual, M. Canalization of the evolutionary trajectory of the human influenza virus. *BMC Biol.* **10,** 38 (2012).
36. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proc. Natl Acad. Sci. USA* **99,** 17209–17214 (2002).
37. Lin, J., Andreasen, V., Casagrandi, R. & Levin, A. S. Traveling waves in a model of influenza A drift. *J. Theor. Biol.* **222,** 437–445 (2003).
38. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science* **305,** 371–376 (2004).

H3N2
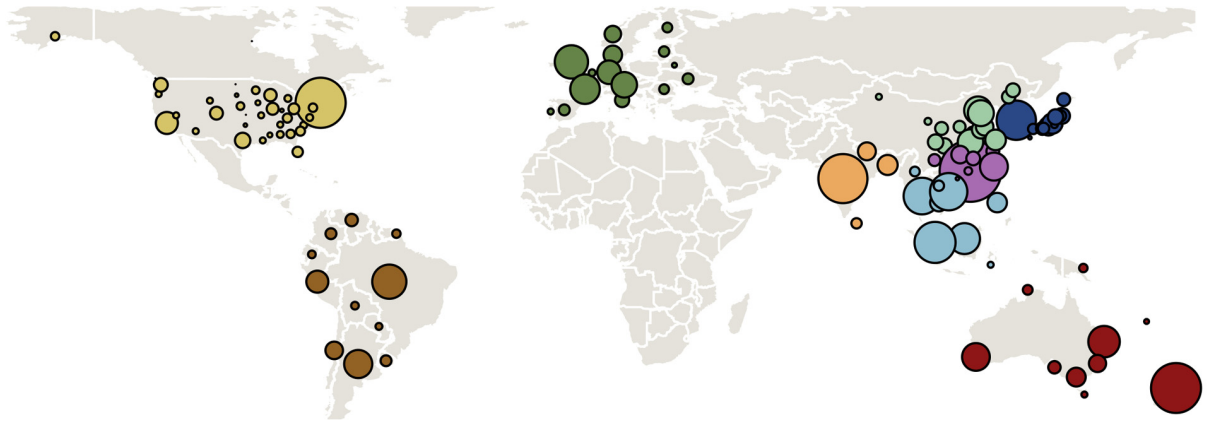
H1N1
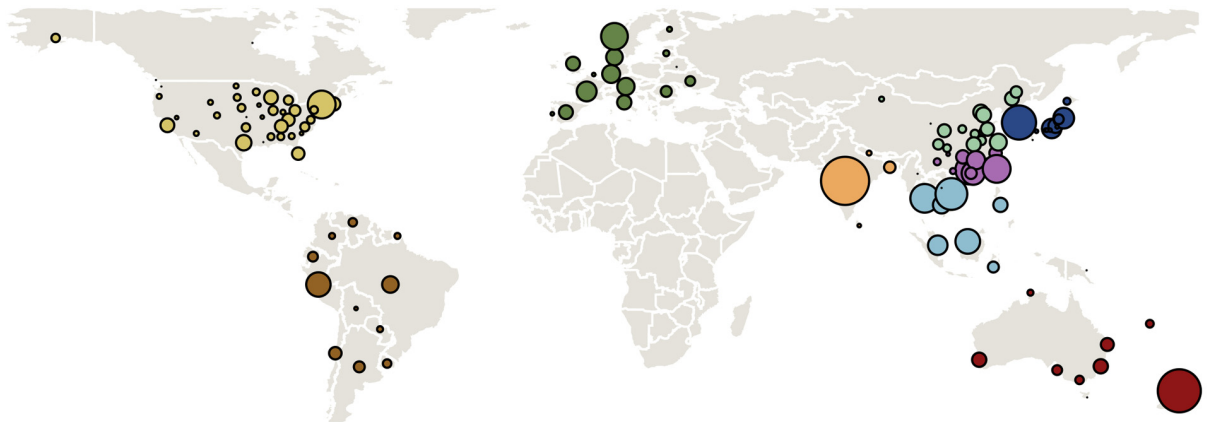
Vic
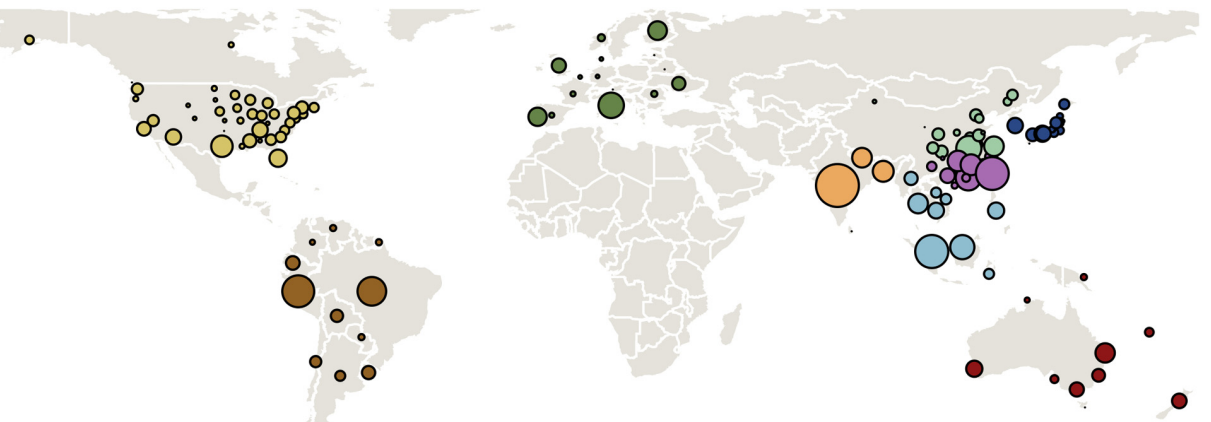
Yam
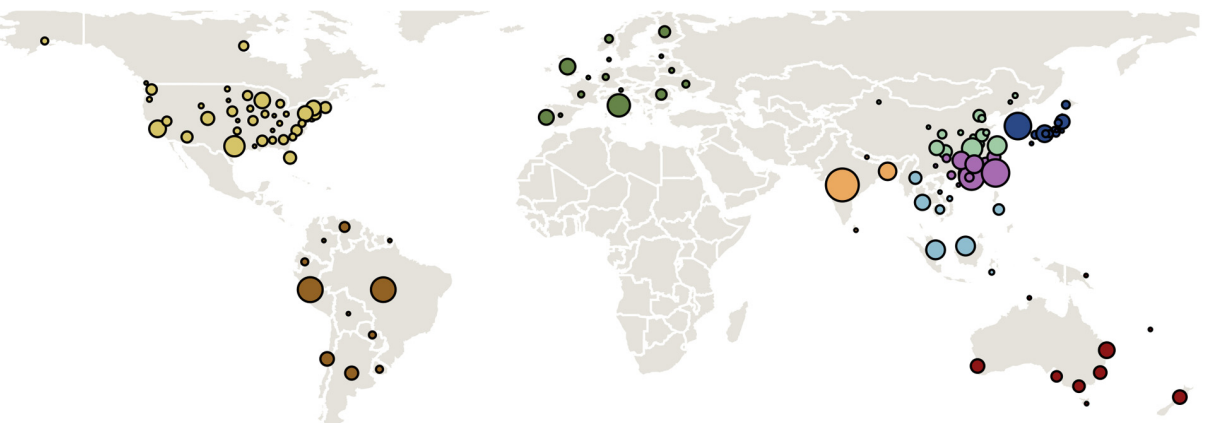
Extended Data Figure 1 | Spatial distribution of 4,006 H3N2, 2,144 H1N1, 1,999 Vic and 1,455 Yam samples. Circle area is proportional to the number of sequenced viruses originating from a location. Colour indicates assignment to one of 9 geographic regions.
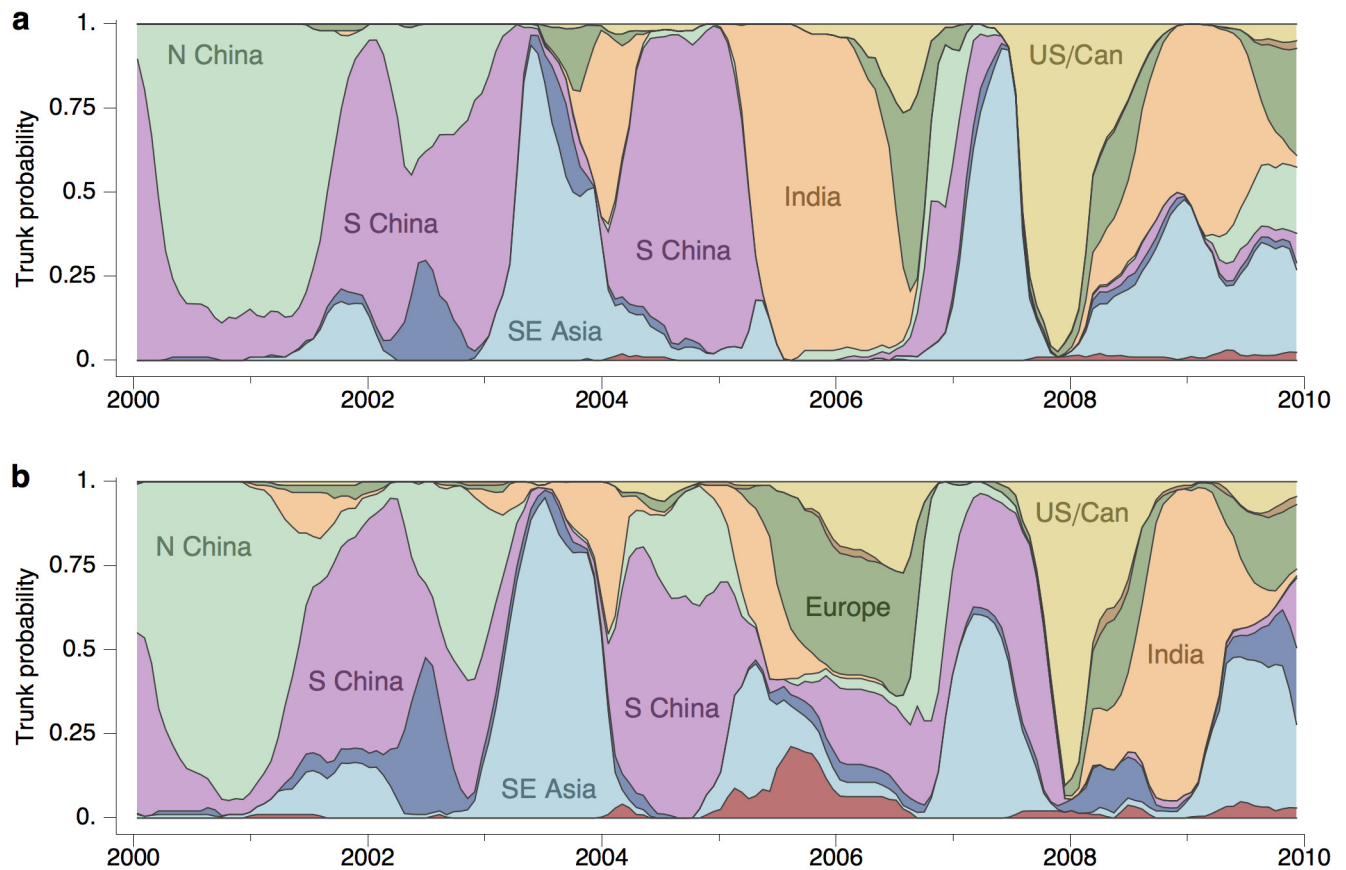
Extended Data Figure 2 | Inferred location of the trunk of H3N2 tree through time in the primary data set (a) and in a smaller secondary data set (b). Coloured width at each time point indicates the posterior support for viruses from a particular geographic location comprising the trunk of the phylogenetic tree. Colours correspond to coloured circles in persistence insets in Fig. 1. The secondary data sets consist of 1,391 H3N2 viruses, 1,372 H1N1 viruses, 1,394 Vic viruses and 1,240 Yam viruses.
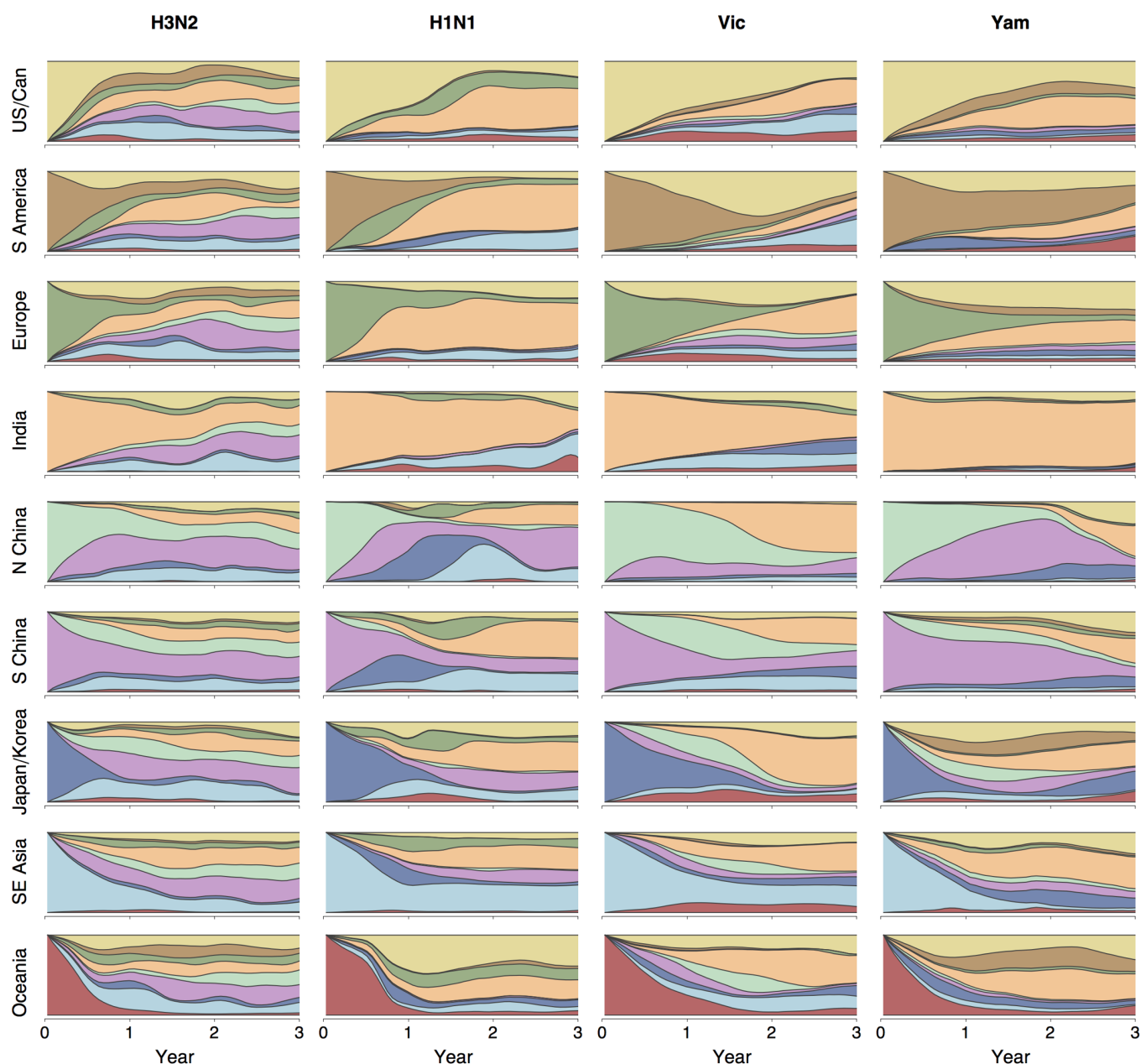
Extended Data Figure 3 | Average inferred geographic history of region-specific samples for H3N2, former seasonal H1N1, Vic and Yam viruses from 2000 to 2012. In each panel, phylogeny tips belonging to a particular region were collected and their phylogeographic histories traced backwards in time averaging across the phylogenetic tree to combine all viruses within each region. The *x*-axis shows number of years backward in time from phylogeny tips from a particular region and the *y*-axis shows the geographic make up as stacked histogram of the ancestors of these tips, where region colour-coding corresponds to the legend in Fig. 1. For example, the top left panel shows the ancestry of USA and Canadian H3N2 viruses. At $x = 0$, all of these viruses are still in the USA or Canada and so an unbroken yellow band takes up the entire *y*. However, at $x = 1$ year, a number of different geographic regions appear on the *y*. This indicates that, 1 year back, ancestors of USA and Canadian viruses are primarily found in Southeast Asia, India and South China. The pattern in the top right panel shows that the ancestors of USA and Canadian Yam viruses more often remain in the USA or Canada with approximately 50% of ancestors remaining 1 year back. Each panel is constructed by averaging across region-specific tips within a tree, but also across sampled posterior trees.
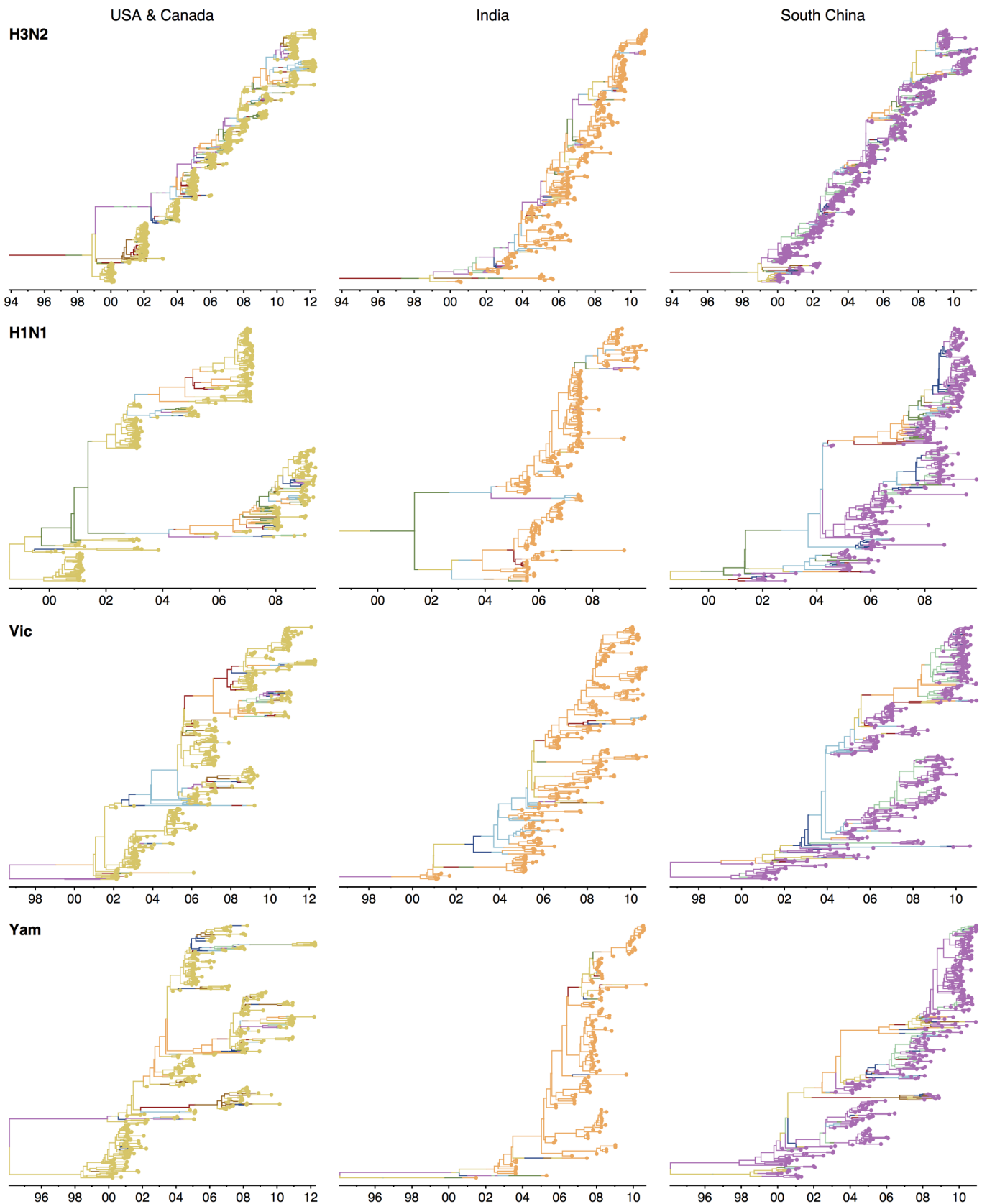
Extended Data Figure 4 | Maximum clade credibility (MCC) trees for region-specific samples from USA/Canada, India and South China for H3N2, H1N1, Vic and Yam viruses. Each tree only contains viruses from a particular geographic region and thus tips are all a single colour within a tree. Branch and trunk colouring have been retained from Fig. 1 to highlight the inferred geographic ancestry of each lineage.
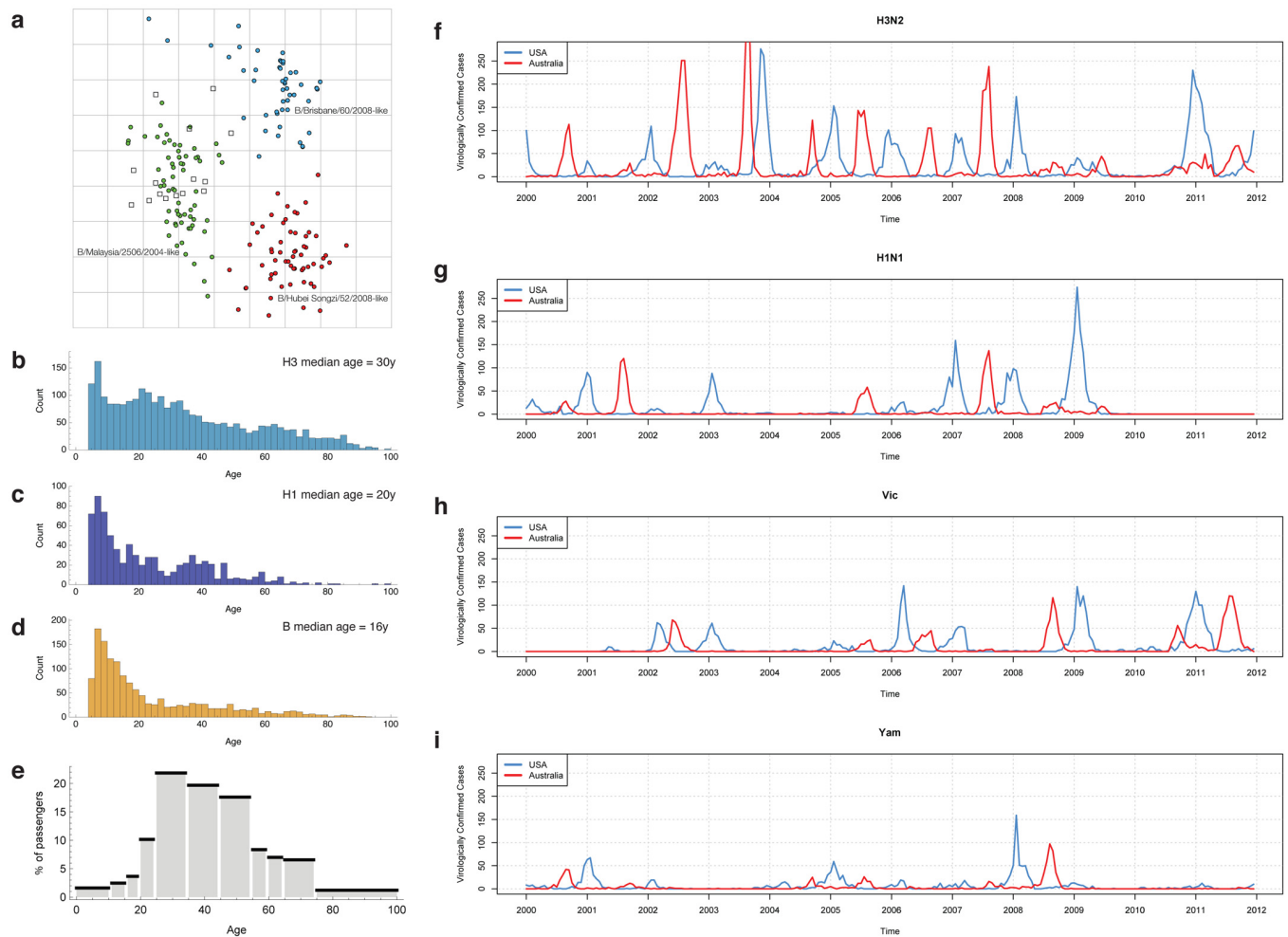
Extended Data Figure 5 | Antigenic map of Vic viruses primarily collected in 2008 (a)**, age distribution of infections for H3N2** (b)**, H1N1** (c) **and B** (d) **in Australia 2000–2011, age distribution of ~102.5 million passengers at London Heathrow and London Gatwick airports during 2011** (e)**, time series of virological characterizations from 2000 to 2012 of viruses from the USA by US CDC and from Australia by VIDRL for H3N2** (f)**, H1N1** (g)**, Vic** (h) **and Yam** (i). In **a**, the positions of strains (coloured circles) and antisera (uncoloured squares) are fit such that the distances between strains and antisera in the map represent the corresponding haemagglutination inhibition (HI) measurements with the least error following Smith *et al.*[38] using data on Vic viruses from the WHO Collaborating Centre for Reference and Research on

Influenza at the Centers for Disease Control and Prevention, Atlanta, Georgia, USA. Strains are coloured by antigenic cluster. Genetic clades corresponding to each antigenic cluster are marked with coloured vertical bars in Fig. 1c. The spacing between grid lines is one unit of antigenic distance corresponding to a twofold dilution of antiserum in the HI assay. In **f–i**, virological characterizations are a surrogate for epidemiological activity that allow for accurate discrimination among H3N2, H1N1, Vic, and Yam viruses. These data generally reflect the relative magnitudes and frequencies of epidemics but in some cases will inflate magnitudes of very small epidemics due to preferential characterization of subtypes circulating at low levels.

a

b
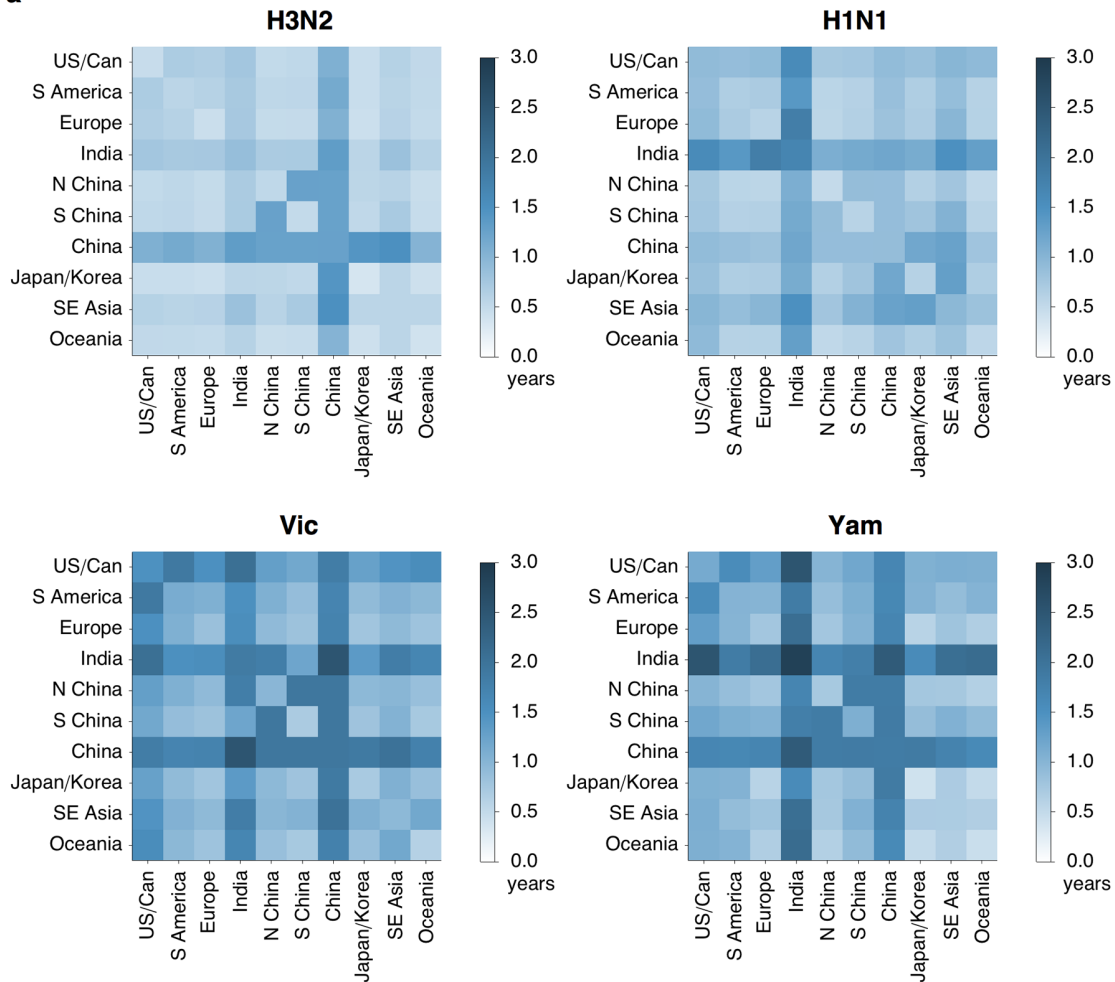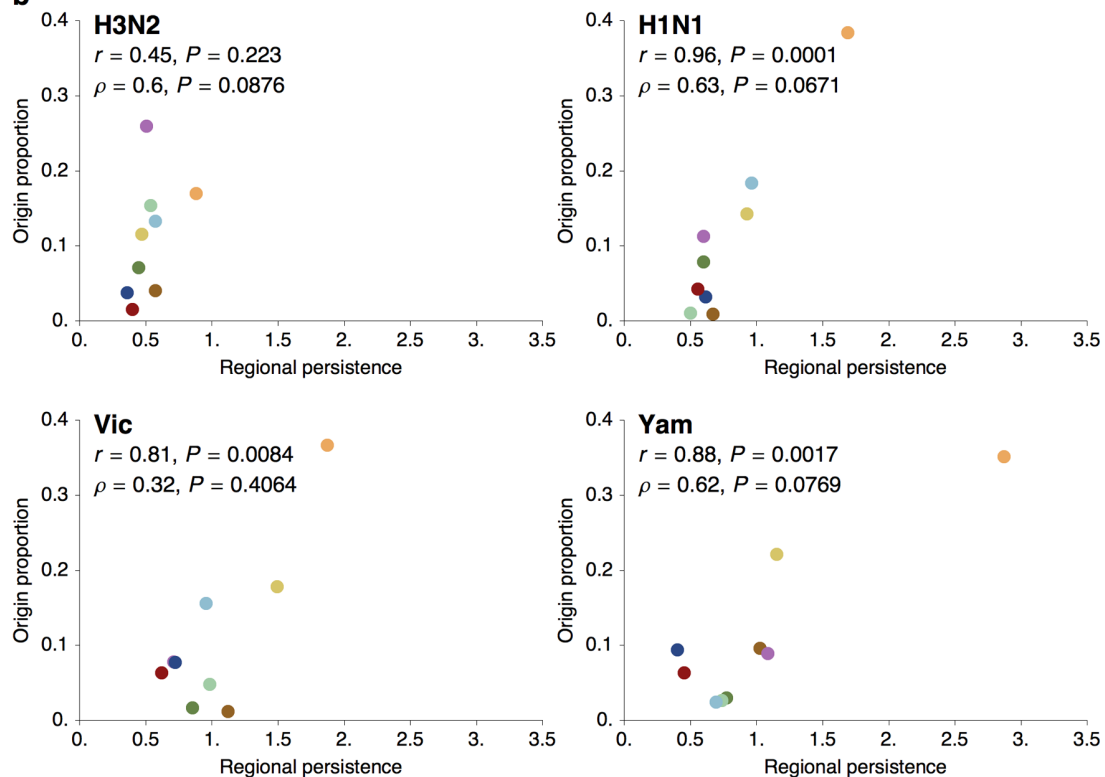
Extended Data Figure 6 | Combined persistence estimates across pairs of regions for H3N2, H1N1, Vic and Yam (a) **and Spearman correlation of a region's persistence vs the region's contribution to phylogenetic ancestry for H3N2, H1N1, Vic and Yam** (b)**.** In **a** and **b**, persistence is measured as the average waiting time in years for a sample to leave its origin backwards in time in the phylogeny, with waiting time averaged across tips within a tree and across sampled posterior trees. In each panel of **a**, the diagonal shows persistence within each of the 9 study regions and within the combined region of 'China', for which nodes in North China and in South China were considered to belong to a single region. The estimates along the diagonal are equivalent to the means shown in Fig. 1. Off-diagonal elements show persistence estimates for pairwise combinations of regions. For example, the off-diagonal for North and South China is exactly equivalent to the diagonal element for 'China' and the off diagonal for 'China' and India represents mean persistence when combining nodes from North China, South China and India. In **b**, origin proportion is measured as the proportion of the time that a region is represented when tracing back 2 or more years from each tip in the phylogeny, averaged across tips within a tree and across sampled posterior trees. Spearman's ρ is not significant for any individual virus. However, the probability of observing 4 instances where each virus has a ρ of at least 0.32 is significant ($P = 0.0017$, bootstrap resampling test).
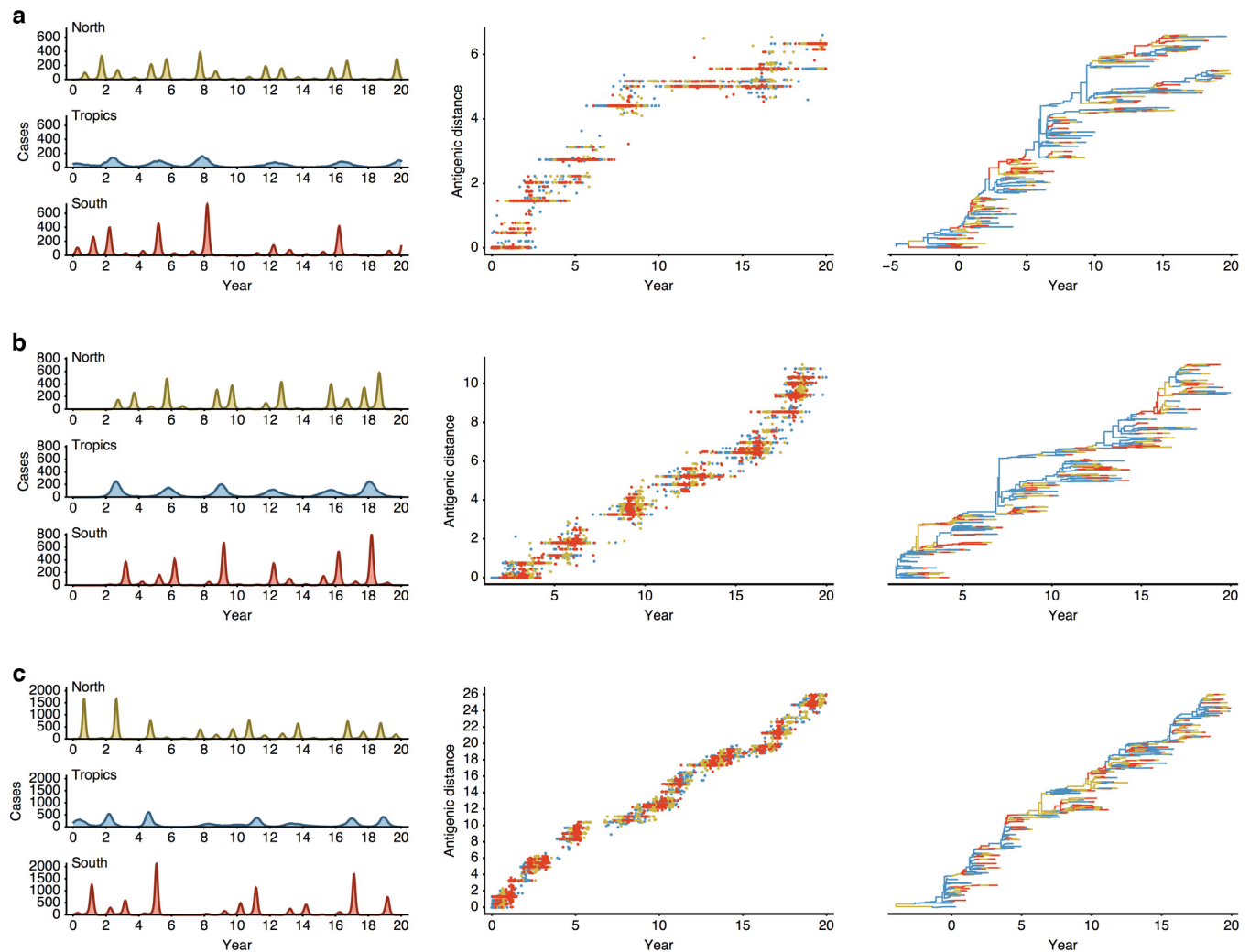
Extended Data Figure 7 | Simulation results for a model parameterized for slow antigenic drift (a)**, moderate antigenic drift** (b)**, and fast antigenic drift** (c)**.** Colours represent geographic regions with tropics in blue, north in yellow and south in red. Region-specific incidence patterns are shown in terms of cases per 100,000 individuals per week, patterns of antigenic drift in terms of increasing antigenic distance (roughly proportional to $\log_2$ HI units) over time and in the geographically labelled phylogeny. The parameterized antigenic mutation rate is 0.00015 antigenic mutations per infection per day in **a**, 0.00035 in **b** and 0.00055 in **c**, while the realized antigenic drift rate is 0.29 antigenic units per year in **a**, 0.58 in **b** and 1.19 in **c**. Between-region mixing is 5.26× faster in adults. Each panel shows output from a single simulation selected from the 112 shown in Fig. 3, and is intended to show model behaviours over a range of parameters, not necessarily the behaviour of particular viruses.
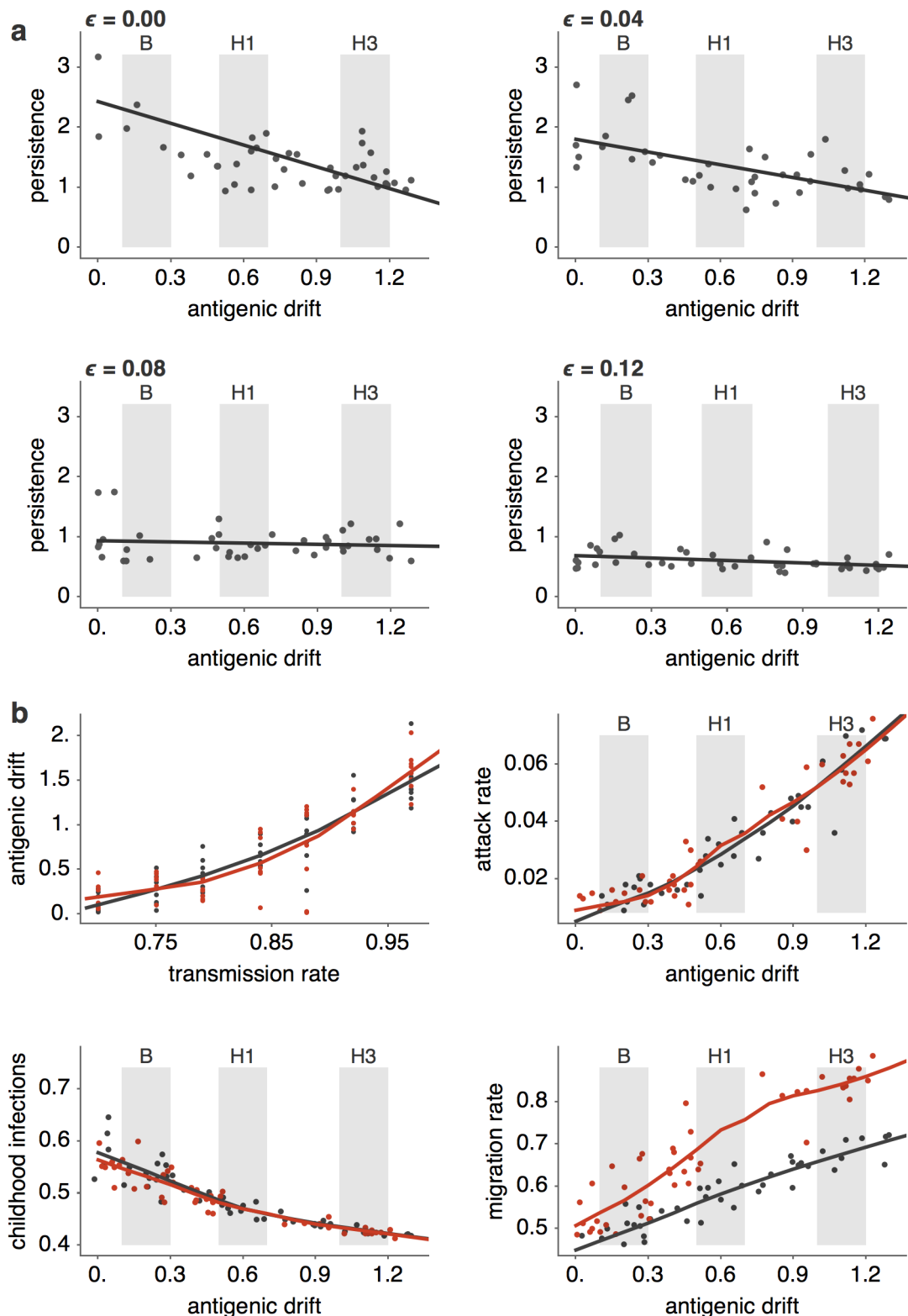
Extended Data Figure 8 | Simulation results showing relationship between antigenic drift and persistence as a function of seasonality (a) and simulation results showing the effects of modulating transmission rate β on model behaviour (b). In **a**, the seasonal forcing parameter ε follows $\varepsilon = 0.00$ (no forcing), $\varepsilon = 0.04$, $\varepsilon = 0.08$ and $\varepsilon = 0.12$ (moderate seasonal forcing). Points represent outcomes from a model in which adults travel between regions at 5.26× the rate of children. Solid black lines represent linear fits to the data. With 4 seasonality scenarios, 7 mutation rates and 8 replicates, there are 224 individual simulations shown. Persistence is measured as the average time in years taken for a tip to leave its region of origin going backwards in time, up the tree. In **b**, transmission rate β in contacts per day is varied and compared to its effect on observed antigenic drift (in antigenic units per year), attack rate per year, proportion of childhood infections and migration rate between regions (in events per viral lineage per year). One antigenic unit is roughly equivalent to one $\log_2$ HI unit. Black points represent outcomes from a model in which children and adults travel between regions at equal rates. Red points represent outcomes from a model in which adults travel between regions at 5.26× the rate of children. Solid black and red lines represent LOESS fits to the data. With 2 travel scenarios, 7 transmission rates and 8 replicates, there are 112 individual simulations shown.

Extended Data Table 1 | Posterior mean estimates (and 95% highest posterior density intervals) across viruses for evolutionary and phylo-geographic parameters

Statistic	H3N2	H1N1	Vic	Yam
Total nucleotide rate*	5.0 (4.8–5.2)	4.4 (4.2–4.6)	2.7 (2.6–2.9)	2.8 (2.6–3.0)
Nonsynonymous rate*	2.2 (2.2–2.3)	1.9 (1.9–2.0)	1.0 (0.9–1.1)	1.0 (0.9–1.0)
Synonymous rate*	2.8 (2.7–2.9)	2.6 (2.5–2.7)	1.8 (1.8–1.9)	1.8 (1.8–1.9)
Antigenic drift rate[†]	1.01 (0.98–1.04)	0.62 (0.56–0.67)	0.42 (0.32–0.52)	0.32 (0.25–0.39)
Diversity[‡]	3.03	4.59	5.46	6.83
TMRCA[§]	3.89	4.53	5.22	7.62
F_{ST}[‖]	0.30	0.36	0.37	0.36
Persistence[¶]	0.50 (0.48–0.54)	0.79 (0.73–0.85)	1.07 (0.98–1.16)	1.03 (0.88–1.21)
Migration rate[#]	1.99 (1.85–2.10)	1.27 (1.18–1.37)	0.93 (0.86–1.02)	0.98 (0.83–1.14)

* Evolutionary rates are measured in terms of 10^{-3} substitutions per site per year.
† Antigenic drift rates are from table 2 in Bedford *et al.*[13], and measures cartographic drift per year in terms of twofold dilution of antiserum in a haemagglutination inhibition (HI) assay.
‡ Diversity of contemporaneous lineages is measured as average time in years for two randomly sampled lineages to share a common ancestor.
§ Time to the most recent common ancestor (TMRCA) of contemporaneous lineages is measured as the average time in years for all lineages to find a common ancestor.
‖ F_{ST} compares diversity within regions to diversity between regions, so that $F_{ST} = (\pi_b - \pi_w)/\pi_b$.
¶ Persistence is calculated as the average number of years for a tip to leave its sampled location, walking backwards up the phylogeny.
# Migration rate is calculated as migration events per lineage per year between any two regions.

Extended Data Table 2 | Posterior mean estimates across viruses and data sets of regional persistence, migration rate and geographic population structure

Statistic	Dataset	H3N2	H1N1	Vic	Yam
Persistence*	Primary§	0.51	0.79	1.07	1.03
Persistence*	Secondary‖	0.53	0.75	1.16	1.11
Persistence*	Alternative¶	0.50	0.76	1.28	1.12
Migration rate†	Primary§	1.96	1.27	0.93	0.97
Migration rate†	Secondary‖	1.89	1.33	0.86	0.90
Migration rate†	Alternative¶	2.00	1.32	0.78	0.89
F_{ST}‡	Primary§	0.30	0.36	0.37	0.36
F_{ST}‡	Secondary‖	0.29	0.35	0.36	0.37
F_{ST}‡	Alternative¶	0.29	0.34	0.36	0.35

* Regional persistence is measured as the average waiting time in years for a sample to leave its origin backwards in time in the phylogeny.

† Migration rate is measured as migration events per lineage per year.

‡ F_{ST} compares diversity within regions to diversity between regions, so that $F_{ST} = (\pi_b - \pi_w)/\pi_b$.

§ The primary data sets consist of 4006 H3N2 viruses, 2144 H1N1 viruses, 1999 Vic viruses and 1455 Yam viruses.

‖ The secondary data sets consist of 1391 H3N2 viruses, 1372 H1N1 viruses, 1394 Vic viruses and 1240 Yam viruses.

¶ The alternative data sets consist of 1967 H3N2 viruses, 1439 H1N1 viruses, 1756 Vic viruses and 1223 Yam viruses divided into 10 geographic regions (USA/Canada, South America, Europe, India, Japan/Korea, Southeast Asia, Oceania, China, Central America and Africa).

LETTER

TH17 cells transdifferentiate into regulatory T cells during resolution of inflammation

Nicola Gagliani[1], Maria Carolina Amezcua Vesely[1]*, Andrea Iseppon[1]*, Leonie Brockmann[2], Hao Xu[1], Noah W. Palm[1], Marcel R. de Zoete[1,3], Paula Licona-Limón[1]†, Ricardo S. Paiva[1], Travers Ching[4], Casey Weaver[5], Xiaoyuan Zi[6]†, Xinghua Pan[7], Rong Fan[6], Lana X. Garmire[4], Matthew J. Cotton[8], Yotam Drier[8], Bradley Bernstein[8], Jens Geginat[9], Brigitta Stockinger[10], Enric Esplugues[11], Samuel Huber[2]§ & Richard A. Flavell[1,3]§

Inflammation is a beneficial host response to infection but can contribute to inflammatory disease if unregulated. The TH17 lineage of T helper (TH) cells can cause severe human inflammatory diseases. These cells exhibit both instability (they can cease to express their signature cytokine, IL-17A)[1] and plasticity (they can start expressing cytokines typical of other lineages)[1,2] upon *in vitro* re-stimulation. However, technical limitations have prevented the transcriptional profiling of pre- and post-conversion TH17 cells *ex vivo* during immune responses. Thus, it is unknown whether TH17 cell plasticity merely reflects change in expression of a few cytokines, or if TH17 cells physiologically undergo global genetic reprogramming driving their conversion from one T helper cell type to another, a process known as transdifferentiation[3,4]. Furthermore, although TH17 cell instability/plasticity has been associated with pathogenicity[1,2,5], it is unknown whether this could present a therapeutic opportunity, whereby formerly pathogenic TH17 cells could adopt an anti-inflammatory fate. Here we used two new fate-mapping mouse models to track TH17 cells during immune responses to show that CD4$^+$ T cells that formerly expressed IL-17A go on to acquire an anti-inflammatory phenotype. The transdifferentiation of TH17 into regulatory T cells was illustrated by a change in their signature transcriptional profile and the acquisition of potent regulatory capacity. Comparisons of the transcriptional profiles of pre- and post-conversion TH17 cells also revealed a role for canonical TGF-β signalling and consequently for the aryl hydrocarbon receptor (AhR) in conversion. Thus, TH17 cells transdifferentiate into regulatory cells, and contribute to the resolution of inflammation. Our data suggest that TH17 cell instability and plasticity is a therapeutic opportunity for inflammatory diseases.

TH17 cells are characterized by secretion of IL-17A, expression of chemokine receptor CCR6 and transcriptional factor RORγt[6]. Their pathogenicity is limited by Foxp3$^+$ T$_{Reg}$ and T regulatory type 1 (TR1) cells[7,8]. Foxp3$^+$ T$_{Reg}$ cells are characterized by the transcription factor Foxp3, whereas TR1 cells secrete high levels of the anti-inflammatory IL-10 and express cell-surface markers CD49b and LAG-3 (refs 7, 9–11). Although TH17, Foxp3$^+$ T$_{Reg}$ and TR1 cells are functionally distinct subsets, they share some features. They are abundant in the intestine, their differentiation is promoted by transforming growth factor β (TGF-β)[12], and both TH17 and TR1 cells express CD49b and high levels of the transcription factor AhR[9,13]. Moreover TH17 cells can transiently co-express RORγt with Foxp3 (refs 14, 15), and IL-17A with IL-10 (refs 10, 16–18).

Despite these similarities, it is unclear if TH17 cells transiently co-express a limited number of genes that are typically associated with regulatory CD4 T cells, or if they can undergo genetic and functional reprogramming resulting in transdifferentiation from one TH type to another.

To track TH17 cell fate towards regulatory states *in vivo*, we crossed IL-17A fate reporter mouse (IL-17A^{CRE} × *Rosa26* STOP$^{fl/fl}$ YFP (R26^{YFP}))[1] with IL-17A^{Katushka} IL-10^{eGFP} Foxp3^{RFP} triple reporter mouse model[9,19]. We call the resulting mouse model Fate$^+$ (Methods, Extended Data Fig. 1a, b) in which, cells that have previously expressed high level of *Il17a,* delete the stop cassette preceding R26^{YFP} and are permanently marked by YFP expression. This enabled us to test if YFP$^+$ cells express IL-17A, IL-10 and Foxp3 *ex vivo* without *in vitro* restimulation.

In steady state TH17 cells are mainly in the small intestine due to the presence of segmented filamentous bacteria (SFB)[12]. Among intestinal CD4 T cells approximately half (48% ±2.7, $n = 18$) of the cells that had expressed IL-17A no longer expressed this cytokine. We call these cells exTH17 cells (IL-17A$^{Katushka-}$ YFP$^+$). Some (4.3% ± 0.3, $n = 18$) intestinal exTH17 cells expressed IL-10^{eGFP}, and some (1% ±0.2, $n = 18$) of them were Foxp3^{RFP} positive (Fig. 1a, b). ExTH17 IL-10^{eGFP+} cells were distinct from TH1, TH2 and TH17 cells since they expressed trace amounts of IFN-γ, were negative for IL-4, and expressed low levels of RORγt and CCR6 respectively (Extended Data Fig. 1c–e). Finally, to test if the presence of TH17 and consequently exTH17 was due to SFB, we treated the mice with vancomycin; both populations were reduced (Fig. 1a, b). Thus under homeostatic conditions, intestinal TH17 cells lose IL-17A expression and a fraction of these exTH17 cells express regulatory features but not characteristic signatures of TH1, TH2 and TH17 cells.

We next analysed TH17 cell plasticity during a self-limiting inflammatory response induced by the injection of anti-CD3 monoclonal antibody[8]. Intestinal TH17 cell expansion was followed by increased exTH17 cells expressing high IL-10^{eGFP} (Fig. 1a, b), although few (2% ±0.2, $n = 8$) of these cells co-expressed IL-10^{eGFP} and Foxp3^{RFP}. The low number of Foxp3$^+$ exTH17 cells prevented, at the time, further studies on these cells.

As exTH17 IL-10^{eGFP+} cells resembled TR1 rather than TH17 cells, we examined them for cell-surface markers that identify TR1 and TH17 cells – LAG-3 (ref. 9), and CCR6 (ref. 12). A high percentage of exTH17 IL-10^{eGFP+} cells were LAG-3 positive but CCR6 negative. Interestingly, in contrast to chronically activated and colitogenic TH17

cells, which are LAG-3 negative[9], TH17 cells expressed low levels of LAG-3 cells during this self-limiting immune response, supporting the idea of an ongoing maturation towards a TR1 cell phenotype. As expected[9] CD49b was equally expressed among the three populations (Extended Data Fig. 2a–c). Like TR1, exTH17 IL-10^{eGFP+} cells expressed low levels of RORγt (Extended Data Fig. 2d) and only trace levels of characteristic TH1 and TH2 genes (Extended Data Fig. 2e). In conclusion, during a self-limiting response, some exTH17 cells resemble TR1 (hereafter named TR1^{exTH17}) rather than TH17 cells.

We next determined TH17 fate during a non-resolving immune response. DNIL-10R transgenic mice have an impairment in IL-10R signalling in CD4 T cells and when treated with anti-CD3 the inflammation does not resolve, but leads to TH17-associated mortality[8]. We found that in Fate DNIL-10R mice, in which immune response cannot be terminated, exTH17 cells tend to acquire a TH1-like phenotype rather than a TR1-like phenotype (Extended Data Fig. 3a–c).

The presence of TR1^{exTH17} cells under steady state conditions suggests two models to explain TR1^{exTH17} cell formation during an immune response. First, TH17 might convert to TR1 cells in steady state and during an immune response, such cells expand (Fig. 1c). Alternatively, TH17 cells might convert to TR1 cells over the course of the response (Fig. 1c). To distinguish these possibilities, we generated a tamoxifen inducible IL-17A^{eGFP} fate mouse model ((iFate)

Methods, Extended Data Fig. 4a–c) in which TH17 cells become YFP$^+$ only after tamoxifen treatment (Fig. 1d and Extended Data Fig. 4d). Through the use of the iFate model we observed that TH17 still convert into TR1 cells, specifically during the immune response (anti-CD3 monoclonal antibody + tamoxifen; Fig. 1d).

We next examined whether TR1^{exTH17} cells undergo transcriptional reprogramming during their conversion from TH17 into TR1 cells. We sequenced the transcriptome of intestinal TR1^{exTH17} cells and compared it to the transcriptomes of bona fide TR1 and TH17 cells isolated from the same mice (Extended Data Fig. 5a, b). As controls, we used exTH17 IL-10$^-$, Foxp3$^+$ T$_{Reg}$ and Foxp3$^+$ Treg IL-10$^+$ cells, again
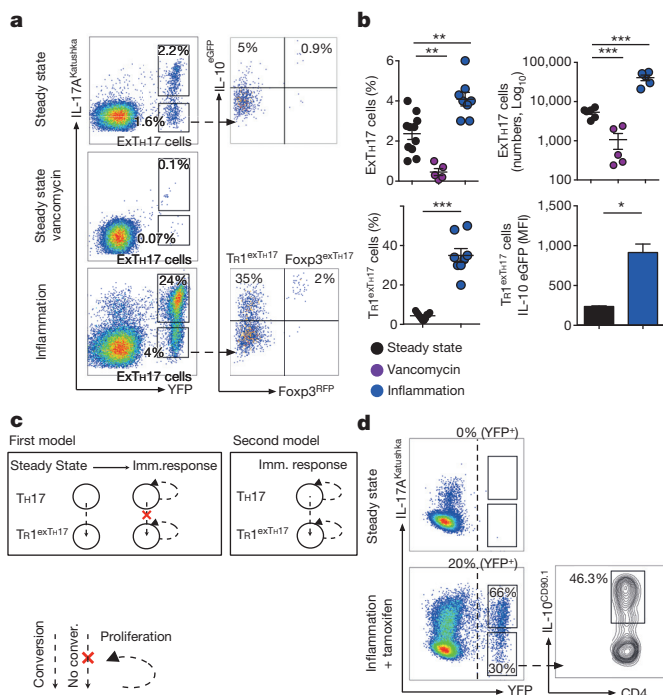
Figure 1 | TH17 cells lose IL-17A and acquire IL-10 in vivo. a, Flow cytometric analysis of small intestinal CD4$^+$ T cells. Steady state, vancomycin-treated, or treated with anti-CD3 monoclonal antibody (inflammation) depicted. **b**, Number and frequencies of exTH17 cells (gated on CD4$^+$ T cells) and TR1^{exTH17} cells (gated on exTH17 cells) are cumulative of two and three independent experiments respectively. IL-10 mean fluorescence intensity (MFI) data ($n = 3$ biological replicates) of one representative experiment out of three are shown. Mean ± s.e.m.; *$P \leq 0.05$, **$P \leq 0.005$, ***$P \leq 0.0005$ by ANOVA (Bonferroni's multiple comparison test) or by t-test for percentage (Mann-Whitney U-test, two tailed) and MFI (paired t-test, two-tailed) of TR1^{exTH17} cells. **c**, Hypotheses: first model, expansion of pre-existing TR1^{exTH17}; second model, conversion of TH17 cells expanded/induced over the course of immune response. **d**, Flow cytometric analysis of small intestinal CD4$^+$ T cells isolated from iFate mice in steady state, and upon anti-CD3 monoclonal antibody and tamoxifen (Inflammation + tamoxifen) treatment. Frequencies of YFP$^+$ cells, TH17 and exTH17 (gated on YFP$^+$ cells), and of IL-10$^+$ cells (gated on exTH17) are representative of two experiments ($n = 6$ biological replicates).
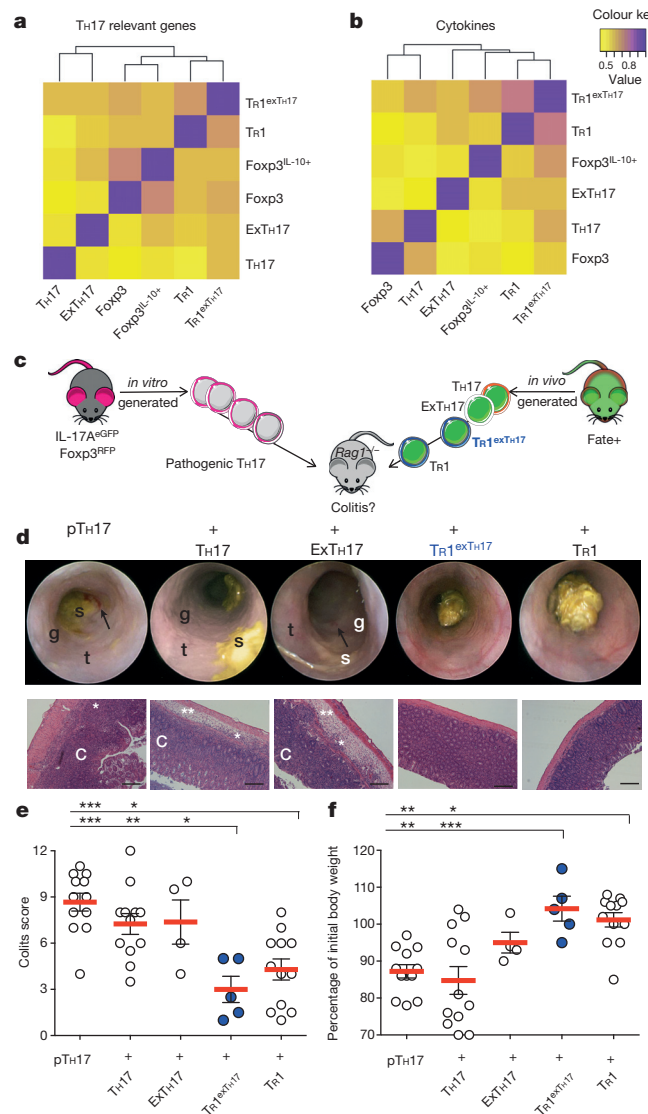
Figure 2 | TR1^{exTH17} cells have a similar gene expression profile and function compare with TR1 cells. a, b, Correlative heatmaps based on the expression of TH17 related genes ($n = 97$) (**a**) and cytokine genes ($n = 191$) (**b**). The indicated cell populations were isolated from the small intestine of 10 anti-CD3 treated Fate$^+$ mice from two independent experiments. **c**, Pathogenic (p)TH17 were differentiated *in vitro* and then injected alone or in combination with the depicted populations into $Rag1^{-/-}$ mice. TH17, exTH17, TR1^{exTH17} and TR1 cells (YFP$^-$) were isolated from the small intestine of Fate$^+$ mice treated with anti-CD3 mAb. **d**, Endoscopic and histological pictures. Scale bars: 200 μm. Endoscopy pictures show stool inconsistency (s), increased mucosal granularity (g), lack of translucency (t) and bleeding (arrow). The histological pictures show oedema (**), inflammation (*) and crypt loss (C). **e, f**, Endoscopic colitis score (**e**) and percentage of initial body weight (**f**). Each dot represents one mouse. Mean and s.e.m. are indicated. *$P \leq 0.05$, **$P \leq 0.005$, ***$P \leq 0.0005$ by ANOVA (Tukey's multiple comparison test).

isolated from the same mice. To determine the relatedness of these subsets, we performed hierarchical clustering based on the expression of literature-search curated TH17-relevant genes[20,21] (Supplementary Table 1). Clustering analyses revealed first that TR1 are a distinct T-cell subset, as different to TH17 or Foxp3[+] T$_{Reg}$ cells as the latter two cell types differ from each other and second that TR1[exTH17] cells cluster together with TR1 rather than with TH17 cells (Fig. 2a). We also performed cytokine-restricted cluster analysis (Supplementary Table 2) showing that TR1 and TR1[exTH17] cluster together (Fig. 2b). Thus, conversion of TH17 into TR1 cells is determined and/or followed by a reprogramming of the TH17-relevant transcriptional profile, a process previously described as transdifferentiation[3,4].

As a final test of whether TR1[exTH17] cells had completed their functional trans-differentiation from TH17 into regulatory TR1 cells, we used a TH17 cell-mediated colitis model[8]. We found that TR1[exTH17] cells had completed their functional reprogramming since they prevented TH17 cell-mediated colitis (Fig. 2c–f).

TH17 cells can also differentiate into TH1-like cells in a TH17 mediated-mouse model for multiple sclerosis (EAE) (Fig. 3a–e)[1,22]. We next wondered whether autoimmune-derived TH17 cells would be able to acquire a regulatory fate. Of note, after anti-CD3 monoclonal antibody treatment, which can block EAE development[17], a fraction of exTH17 cells acquired IL-10, not IFN-γ (Fig. 3a–e). Thus, some exTH17 cells, which developed during an autoimmune response, can still convert into TR1 cells.

Furthermore, encephalitogenic TH17 cells can be recruited to the small intestine after anti-CD3 monoclonal antibody treatment[17]. In our current studies some intestinal T cells specific for the disease-driving antigen of EAE (myelin oligodendrocyte glycoprotein

(MOG)) were TR1[exTH17] cells. We also observed that MOG-specific exTH17 cells express IL-10 to a greater extent than non-MOG-specific exTH17 cells (Fig. 3f–i). Likewise, in iFate mice TH17 cells labelled during EAE onset converted to TR1 cells (Extended Data Fig. 6a, b). Thus pathogenic autoantigen-specific TH17 cells can convert to TR1[exTH17] cells.

We next addressed if TH17 cells convert to TR1 cells during an immune response, which physiologically promotes host tolerance to infection[23]. *Nippostrongylus brasiliensis* infection elicits a type-2 immune response that drives worm expulsion[24]. In addition to TH2 cells, TH17 and TR1 also expand in response to *N. brasiliensis*[9] and while IL-17 contributes to tissue damage, IL-10 prevents tissue damage[24]. We therefore asked if TH17 cells convert to TH2 cells[25] or to TR1 cells during *N. brasiliensis* infection. During the primary immune response to *N. brasiliensis*, TH17 cells lost IL-17A expression and some showed a TH2 phenotype[25]. However, when we re-infected the mice with *N. brasiliensis* we observed TH17 conversion into a TR1 cell-phenotype (Fig. 3j–l). We confirmed these findings in iFate mice (Extended Data Fig. 6c–e). Thus, TH17 cells can become TR1 cells during the secondary response to *N. brasiliensis* infection and this may limit potentially destructive type 1 immune responses.

Finally we asked if TH17 conversion to TR1 cells occurs in response to acute bacterial infection. *S. aureus* causes sepsis in human and patients with TH17 associated gene deficiency suffer recurrent *S. aureus* infections[26]. Intravenously *S. aureus* infected Fate[+] mice (Extended Data Fig. 7a–c) and iFate mice (Extended Data Fig. 7d, e) accumulated TH17 cells in the small intestine, as we previously showed[17], and in both models TH17 cells acquired a TR1 cell phenotype.
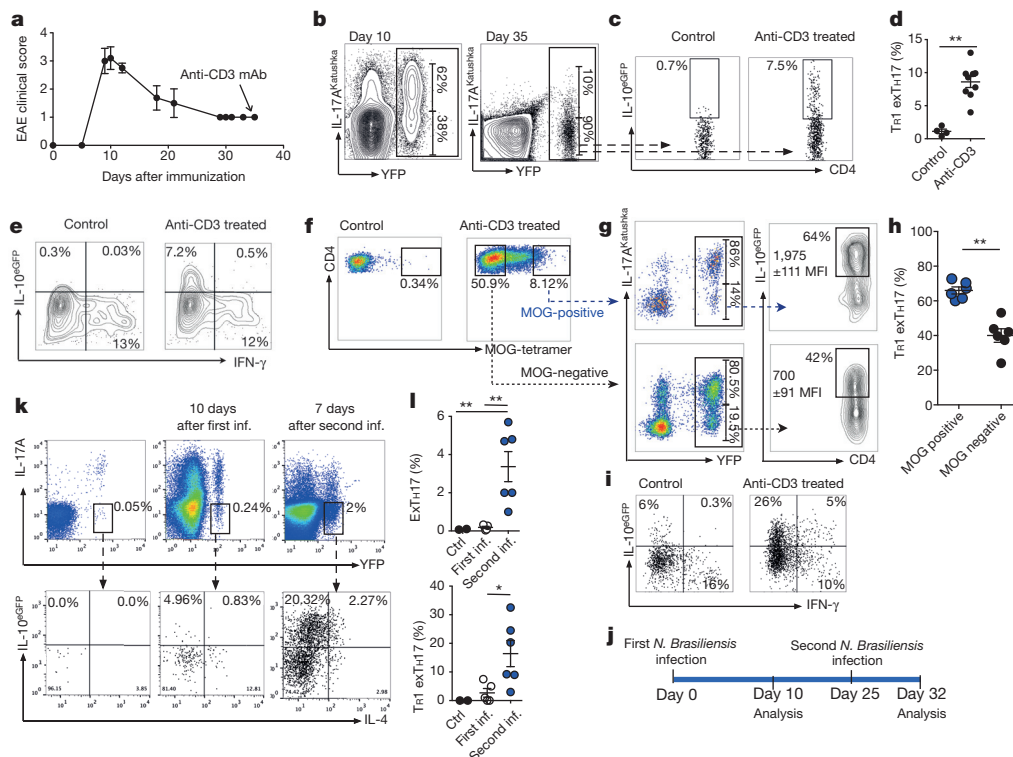
Figure 3 | TR1[exTH17] cell development in EAE and helminth infection.
a, Clinical EAE-score. Anti-CD3 was injected 35 days after MOG-immunization. **b,** Flow cytometric analysis of TH17 and exTH17 (gated on YFP[+] cells) cells in dLNs. **c, d,** Flow cytometric analysis (**c**) and percentages of TR1[exTH17] cells (gated on exTH17) in dLNs (**d**). **e,** IFN-γ/IL-10[eGFP] expression of exTH17 cells. **a–e,** Representative of three independent experiments. **f,** Flow cytometric analysis of MOG-tetramer staining of intestinal CD4 cells of EAE mice left untreated (control) or treated with anti-CD3. **g, h,** Representative flow cytometric analysis (**g**) and frequencies of TR1[exTH17] (gated on MOG[±] cells)

(**h**). Each dot represents one mouse. IL-10 MFIs (average of three mice ± s.e.m.) are reported. **P ≤ 0.05 by Mann–Whitney U-test, two tailed. **i,** IFN-γ/IL-10[eGFP] expression of exTH17 isolated from small intestine of EAE mice. **f–i,** Representative of two independent experiments. **j,** Schematic of the experiment. **k, l,** Flow cytometric analysis and frequencies of exTH17 and TR1[exTH17] from the lung (upper panel: gated on CD4[+] T cells; lower panel: gated on exTH17 cells). One experiment of three is shown. Mean ± s.e.m., *P ≤ 0.05, **P ≤ 0.005 by ANOVA (Tukey's multiple comparison test).
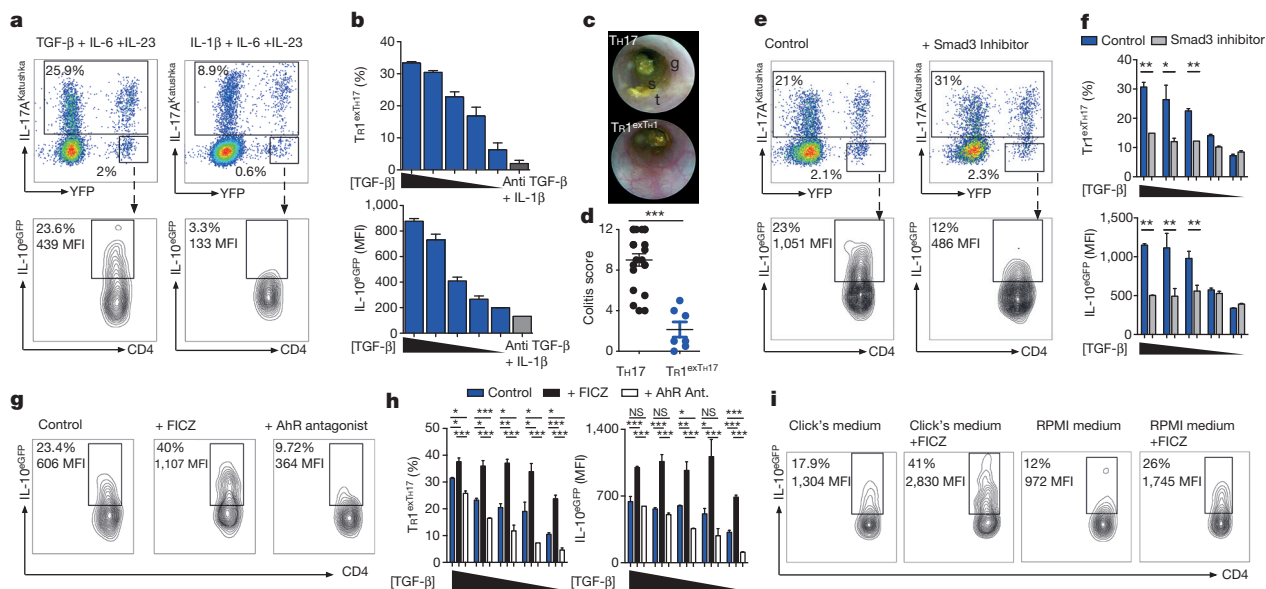
Figure 4 | TGF-β1 via Smad3, and AhR support the conversion of TH17 to TR1. TH17 cells were differentiated *in vitro* in the presence of IL-6, IL-23 with TGF-β1 or with IL-1β and anti-TGF-β monoclonal antibody. TGF-β1 was diluted 1:2 starting from 4 ng ml⁻¹. **a**, Flow cytometric analysis of TH17, exTH17 and TR1^exTH17 (gated on exTH17). **b**, Percentages and IL-10 MFI of TR1^exTH17 cells. Technical replicates (*n* = 2) of one experiment out of seven. **c, d**, Endoscopic pictures and score of mice injected with TH17 or TR1^exTH17 cells polarized with TGF-β1. Stool inconsistency (s), increased mucosal granularity (g) and a lack of translucency (t). Each dot denotes one biological replicate. Mean and s.e.m., ***P ≤ 0.0005 by Mann Whitney *U*-test, two tailed. **e**, Flow cytometric analysis of TH17, exTH17 and TR1^exTH17 cells cultured in the presence of TGF-β1 (diluted as above), IL-6, IL-23 ± Smad3 inhibitor. **f**, Percentages and IL-10 MFI of TR1^exTH17 cells. Technical replicates (*n* = 3) of one experiment out of five are shown. Mean and s.e.m., *P ≤ 0.05, **P ≤ 0.005 by paired *t*-test. **g**, Flow cytometric analysis of TR1^exTH17 (gated on exTH17) cultured in the presence of TGF-β1 (diluted as above), IL-6, IL-23 ± AhR ligand (FICZ) or AhR antagonist. **h**, Percentages and IL-10 MFI of TR1^exTH17 (gated on exTH17). Technical replicates (*n* = 3) of one experiment out of five are shown. Mean and s.e.m.; *P ≤ 0.005, **P ≤ 0.005, ***P ≤ 0.0005 by ANOVA (Tukey's multiple comparison test). NS, non-significant. **i**, Flow cytometric analysis of TR1^exTH17 cells cultured in the presence of TGF-β1, IL-6, IL-23 ± FICZ in the indicated medias. One experiment of two is shown.

Thus, conversion of TH17 into TR1 cells is a physiological mechanism, occurring in steady state and favoured during worm and bacterial infection.

We sought candidate pathways that drive TH17 conversion into TR1 cells. Expression of twelve genes from our list of TH17-relevant genes appeared to be higher in both TR1 and TR1^exTH17 cells, than in TH17 cells with 9 of them being associated with TGF-β signalling (Extended Data Fig. 8a, b). TGF-β1, in combination with IL-6 and IL-23, promotes *in vitro* development of potentially pathogenic TH17 cells[16,27]. Some of these TH17 cells expressed IL-10 (ref. 16). We found that TGF-β1, in contrast to IL-1β, which is known to induce IL-10^Negative TH17 cells[18,28], promoted TH17 cell plasticity and conversion to TR1 cells in a dose-dependent manner (Fig. 4a, b and Extended Data Fig. 9a). Since *in vivo* T cells are likely exposed to both TGF-β1 and IL-1β simultaneously we tested TR1^exTH17 cell development in the presence of both cytokines and found that TR1^exTH17 cells mature normally. However neutralizing TGF-β monoclonal antibody impaired the development of TR1^exTH17 cells (Extended Data Fig. 9b). Importantly, while TH17 cells generated with TGF-β1/IL-6/IL-23 are able to promote colitis, TR1^exTH17 cells generated under the same conditions failed to induce disease. Thus, TGF-β1 is important for TR1^exTH17 development and despite TGF-β1, TH17 cells remain colitogenic as long as they do not convert into TR1 cells (Fig. 4c, d).

Among the TGF-β signalling pathway molecules, Smad3 decreases RORγt activity and therefore reduces TH17 cell development[29]. We asked if TGF-β1 promotes TH17 to TR1 conversion by modulating Smad3. Thus, when we blocked Smad3 during *in vitro* TH17 differentiation, induction of TR1^exTH17 cells was reduced (Fig. 4e, f and Extended Data Fig. 9c). Thus, TGF-β1, likely through Smad3, promotes TH17 to TR1 conversion.

TH17 and TR1 cells differentiated in the presence of TGF-β1 express a high level of AhR[13,28]. AhR promotes *Il10* transactivation and TR1

generation is defective in mutant Ahr^d mice[13]. Thus, the role of Ahr expression in potentially inflammatory TH17 cells could be to enable a switch to a regulatory fate and terminate the immune response[30]. Indeed, as reported[28] CD4 T cells, skewed towards TH17 in the presence of TGF-β1, express high levels of *AHR* (Extended Data Fig. 9d). Finally, to test whether AhR activation influences TH17 to TR1^exTH17 conversion we added an AhR ligand (FICZ) or AhR antagonist to the culture. FICZ significantly enhanced the development of TR1^exTH17 cells while the AhR antagonist reduced the conversion (Fig. 4g, h). Moreover, when replacing Click's medium, which is rich in AhR ligands, with RPMI, poor in AhR ligands, the development of TR1^exTH17 was reduced. Adding FICZ to RPMI medium rescued the conversion (Fig. 4i). These *in vitro* generated TR1^exTH17 cells also exhibited regulatory function (Extended Data Fig. 9e). Finally we purified intestinal TH17 cells and showed that some of these cells when re-stimulated *in vitro* with TGF-β1 and FICZ convert into TR1 cells (Extended Data Fig. 9f).

Overall our study shows TH17 cells can transdifferentiate to TR1 cells during an immune response and in the presence of TGF-β1, AhR activation promotes this conversion. We believe that TH17 cell plasticity might be exploited to develop new and more effective therapies that restore immune tolerance in chronic inflammatory/autoimmune diseases without incurring the deleterious side-effects associated with current systemic immunosuppressive therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Hirota, K. *et al.* Fate mapping of IL-17-producing T cells in inflammatory responses. *Nature Immunol.* **12,** 255–263 (2011).

2. Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J. Exp. Med.* **204,** 1849–1861 (2007).
3. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462,** 587–594 (2009).
4. Komatsu, N. *et al.* Heterogeneity of natural Foxp3$^+$ T cells: a committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proc. Natl Acad. Sci. USA* **106,** 1903–1908 (2009).
5. Nylander, A. & Hafler, D. A. Multiple sclerosis. *J. Clin. Invest.* **122,** 1180–1188 (2012).
6. Huber, S., Gagliani, N. & Flavell, R. A. Life, death, and miracles: Th17 cells in the intestine. *Eur. J. Immunol.* **42,** 2238–2245 (2012).
7. Roncarolo, M. G. & Battaglia, M. Regulatory T-cell immunotherapy for tolerance to self antigens and alloantigens in humans. *Nature Rev. Immunol.* **7,** 585–598 (2007).
8. Huber, S. *et al.* Th17 cells express interleukin-10 receptor and are controlled by Foxp3$^-$ and Foxp3$^+$ regulatory CD4$^+$ T cells in an interleukin-10-dependent manner. *Immunity* **34,** 554–565 (2011).
9. Gagliani, N. *et al.* Coexpression of CD49b and LAG-3 identifies human and mouse T regulatory type 1 cells. *Nature Med.* **19,** 739–746 (2013).
10. Heinemann, C. *et al.* IL-27 and IL-12 oppose pro-inflammatory IL-23 in CD4+ T cells by inducing Blimp1. *Nature Commun.* **5,** 3770 (2014).
11. Okamura, T. *et al.* TGF-β3-expressing CD4$^+$CD25$^-$LAG3$^+$ regulatory T cells control humoral immune responses. *Nature Commun.* **6,** 6329 (2015).
12. Littman, D. R. & Rudensky, A. Y. Th17 and regulatory T cells in mediating and restraining inflammation. *Cell* **140,** 845–858 (2010).
13. Apetoh, L. *et al.* The aryl hydrocarbon receptor interacts with c-Maf to promote the differentiation of type 1 regulatory T cells induced by IL-27. *Nature Immunol.* **11,** 854–861 (2010).
14. Zhou, L. *et al.* TGF-β-induced Foxp3 inhibits T$_H$17 cell differentiation by antagonizing RORγt function. *Nature* **453,** 236–240 (2008).
15. Beriou, G. *et al.* IL-17-producing human peripheral regulatory T cells retain suppressive function. *Blood* **113,** 4240–4249 (2009).
16. McGeachy, M. J. *et al.* TGF-β and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain T$_H$-17 cell-mediated pathology. *Nature Immunol.* **8,** 1390–1397 (2007).
17. Esplugues, E. *et al.* Control of T$_H$17 cells occurs in the small intestine. *Nature* **475,** 514–518 (2011).
18. Zielinski, C. E. *et al.* Pathogen-induced human T$_H$17 cells produce IFN-γ or IL-10 and are regulated by IL-1β. *Nature* **484,** 514–518 (2012).
19. Kamanaka, M. *et al.* Expression of interleukin-10 in intestinal lymphocytes detected by an interleukin-10 reporter knockin tiger mouse. *Immunity* **25,** 941–952 (2006).
20. Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151,** 289–303 (2014).
21. Yosef, N. *et al.* Dynamic regulatory network controlling TH17 cell differentiation. *Nature* **496,** 461–468 (2014).
22. Cua, D. J. *et al.* Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature* **421,** 744–748 (2003).
23. Gause, W. C., Wynn, T. A. & Allen, J. E. Type 2 immunity and wound healing: evolutionary refinement of adaptive immunity by helminths. *Nature Rev. Immunol.* **13,** 607–614 (2013).
24. Chen, F. *et al.* An essential role for T$_H$2-type responses in limiting acute tissue damage during experimental helminth infection. *Nature Med.* **18,** 260–266 (2012).
25. Panzer, M. *et al.* Rapid *in vivo* conversion of effector T cells into Th2 cells during helminth infection. *J. Immunol.* **188,** 615–623 (2011).
26. Milner, J. D. *et al.* Impaired T$_H$17 cell differentiation in subjects with autosomal dominant hyper-IgE syndrome. *Nature* **452,** 773–776 (2008).
27. Lee, Y. *et al.* Induction and molecular signature of pathogenic T$_H$17 cells. *Nature Immunol.* **13,** 991–999 (2012).
28. Ghoreschi, K. *et al.* Generation of pathogenic T$_H$17 cells in the absence of TGF-beta signalling. *Nature* **467,** 967–971 (2010).
29. Martinez, G. J. *et al.* Smad3 differentially regulates the induction of regulatory and inflammatory T cell differentiation. *J. Biol. Chem.* **284,** 35283–35286 (2009).
30. Stockinger, B., Di Meglio, P., Gialitakis, M. & Duarte, J. H. The aryl hydrocarbon receptor: multitasking in the immune system. *Annu. Rev. Immunol.* **32,** 403–432 (2014).

METHODS

Mice. C57BL/6 *Rag1*$^{-/-}$ and Rosa26^{flox}STOP^{flox} eYFP mice were purchased from The Jackson Laboratories. IL-17A^{CRE} and IL-10$^{Th1.1}$ (CD90.1) mice were kindly provided from B. Stockinger and C. Weaver respectively[1,31]. IL-17-IGCE iFate mice, Foxp3^{RFP}, IL-10^{eGFP}, IL-17A^{Katushka}, IFN-$\gamma^{Katushka}$ reporter mice and CD4dnIL-10Rα (DNIL-10R) were generated and breed in our laboratory. reporter mice were purchased from Jackson Laboratory. The Fate$^+$ mice results from the breeding of the original Foxp3^{RFP} IL-10^{eGFP} IL-17A^{Kata} mice[9,17,19] with IL-17A^{CRE} R26^{YFP} mice[1]. In this model, only high level of *Il17a* transcription induces the expression of Cre recombinase, which deletes the stop sequence 5′ to YFP. In this mouse, cells that have previously expressed high level of *Il17a*, delete the stop cassette and are thus permanently marked by the expression of YFP. Importantly it has been previously described that this IL-17A fate reporter allele faithfully marks T$_H$17 cells that have acquired full effector function[1]. Of note, we observed that all IL-17A bright cells are YFP$^+$, whereas IL-17A dim cells remain YFP negative, confirming the data already published whereby only IL-17A high expressing cells, fully differentiated T$_H$17 cells, are permanently marked with YFP[1].

To generate the DNIL-10R FATE, the CD4-DNIL-10R were crossed with IL-17A^{CRE} R26^{YFP} IL-10^{eGFP} Foxp3^{RFP}.

All mice were kept under specific pathogen-free (SPF) conditions in the animal facility at Yale University. We used age- and sex-matched littermates between 12 and 20 weeks of age. Animal procedures were approved by the Institutional Animal Care and Use Committee of Yale University. Both female and male mice were used in experiments. Wherever possible, preliminary experiments were performed to determine requirements for sample size, taking into account resources available and ethical use. Exclusion criteria such as inadequate staining or low cell yield due to technical problems were pre-determined. Animals were assigned randomly to experimental groups. Each cage contained animals of all the different experimental groups expected for the mice treated with antibiotic.

Generation of inducible Fate mice. The IL-17A IRES-eGFP-CRE-ERT2 mice (iFate) mice were generated following the same targeting strategy used previously to generate the IL-17A^{eGFP} mice[17]. Briefly, a cassette encoding for a fusion protein consisting of the Internal Ribosome Entry Site (IRES), eGFP, Cre and human modified Oestrogen Receptor (ERT2) (IGCE) was linked to a Frt-flanked neomycin (NEO) encoding cassette. The IGCE-NEO construct was cloned into a plasmid containing two homology arms on the *Il17a* gene: the 5′ homology arm corresponds to a 4,445-bp fragment of the *Il17a* gene to the fourteenth base pairs after the stop codon of the gene using an Asc cloning site while the 3′ homology arm of the targeting construct consists of the genomic sequence of 3,258 bp spacing from fifteenth base pairs after the stop codon of the *Il17a* gene using a Not cloning site.

Drug-resistant ES cell clones were screened for homologous recombination by PCR. To obtain chimeric mice, correctly targeted ES clones were injected into C57BL/6 blastocysts, which were then implanted into CD1 pseudopregnant foster mothers. Male chimaeras were bred with C57BL/6 to screen for germline transmitted offspring. Germline transmitted mice were bred with germline Flippase expressing transgenic mice to remove the neomycin gene.

Mice bearing the construct were screened by PCR and bred with germline Flp expressing transgenic mice to remove the neomycin gene. After removal of the NEO cassette, IL-17A iFate mice were crossed with R26^{eYFP} and all of the cells that actively express IL-17A were eGFP$^+$ but still YFP$^+$ negative. However, unlike in Fate$^+$ mice, IL-17A expressing cells become permanently marked as YFP$^+$ after treatment with tamoxifen as ERT2 sequesters the Cre in the cytoplasm until tamoxifen binds to ERT2.

The efficiency of CRE-mediated recombination after tamoxifen is reported as frequencies of YFP$^+$ cells (YFP) in Fig. 1d and Extended Data Figs 4, 6 and 7 and is the result of the following calculation: YFP$^+$ cells (Gate2+3) / (IL-17A^{eGFP+} YFP$^-$ cells (Gate 1) + YFP$^+$ cells(Gate2+3).

Anti-CD3, antibiotic and tamoxifen treatments. Mice were injected with anti-CD3 mAb (2C11,15–50 μg per mouse) intra-peritoneally two times every other day. Usually the mice were sacrificed 4 h after the last injection, unless differently indicated. Vancomycin was dissolved in water to a final concentration of 0.5 g l^{-1} and administered in drinking water for 4 weeks before starting the experiment. Tamoxifen (Sigma) was dissolved in corn oil (Fluka, Sigma) to a final concentration of 20 mg ml^{-1}. Mice were injected with tamoxifen (4 mg each) one day before each anti-CD3 monoclonal antibody injection or as depicted in the scheme of the other experiments. To avoid to interfere with the effect of anti-CD3 mAb and/or the migration of cells to the intestine, the oil + tamoxifen was injected subcutaneously. Only for the experiments involving *N. brasiliensis* infection we injected tamoxifen intraperitoneally (i.p.).

Lymphocyte isolation from small intestine. After removal of the Peyer's patches, we isolated intraepithelial lymphocytes (IELs) and lamina propria lymphocytes (LPLs) by incubation with 1 mM DTE at 37 °C for 30 min (for IEL), followed by further digestion with collagenase from Clostridium Histolyticum (#2139 SIGMA) and DNase at 37 °C for 1 h (for LPL). We then further separated cells with a Percoll gradient. Unless otherwise indicated, we isolated cells from the small intestine (duodenum, ileum and jejunum) of mice treated with antibodies to CD3.

Flow cytometry antibodies and intracellular cytokine staining. We stained mouse T cells with monoclonal antibodies to CD4 (GK1.5, Cat # 100428 or RM4-5 Cat # 100536), CD8 (53-6.7, Cat # 100722), NK1.1 (RM4-5 Cat # 100536), CD19 (6D5 Cat # 115508), CD11b (M1/70 Cat # 101216), CD11c (N418 Cat # 117318), $\gamma\delta$TCR (GL3 Cat # 118123), CD210 (BD Bioscience, Cat # 559914), LAG-3 (C9B7W Cat # 125209), CD49b (HMa2 Cat # 103506) and CCR6 (29-2L17 Cat # 129817), all antibodies expected where indicated are purchased from eBiolegend. Importantly, CD49b and LAG-3 staining were performed at 37 °C for 45 min. Although in the figure legends we referred only to CD4$^+$ T cells, in each FACS related experiment and FACS-sorting experiment we have specifically analysed CD4$^+$ T cells CD8$^-$, NK1.1$^-$, CD19$^-$, CD11b$^-$, CD11c$^-$, $\gamma\delta$TCR$^-$. For intracellular cytokine staining the cells were re-stimulated for 3 h at 37 °C with phorbol 12-myristate 13-acetate (PMA) (Sigma, 50 ng ml^{-1}) and ionomycin (Sigma, 1 μg ml^{-1}) in the presence of Golgistop (BD Bioscience). Cells were then fixed in paraformaldehyde for 20 min at room temperature. After washing, cells have been permeabilized (NP40) and stained at 4 °C with anti-IL-17A (TC11-18H10.1 Cat # 506925), anti-IFNγ (BD Bioscience, Cat # 554412), anti-IL-4 (BD Bioscience, Cat # 554435) and anti-Rorγt (BD Bioscience Cat # 553178) antibodies for 30 min. Lymphocytes were re-suspended in PBS, 0.5% FBS, 5 mM EDTA and acquired with an LSRII cytometer (BD Bioscience).

***In vitro* T$_R$1^{exT_H17} cell differentiation.** We FACS sorted CD4$^+$ Foxp3^{RFP-} IL-17A$^{Katushka-}$ IL-10^{eGFP-} R26^{YFP-} cells with FACSAria II Cell Sorter (BD Biosciences) and activated them with plate-bound monoclonal antibodies to CD3 (10 μg ml^{-1}, 145-2C11) and CD28 (1–2 μg ml^{-1}, PV-1) in the presence of mouse recombinant TGF-β (0.25–4 ng ml^{-1}), IL-6 (20 ng ml^{-1}), IL-23 (20 ng ml^{-1}), and antibodies to IFN-γ (XMG1.2, 10 μg ml^{-1}) and IL-4 (11B11, 10 μg ml^{-1}). When specified, IL-1β (50 ng ml^{-1}), antibodies to TGF-β (5 μg ml^{-1}, 1D11,) and FICZ (100 nM; Enzo Life Sciences), Smad3 inhibitor (SIS3, 3 μM, EDM Millipore)[32] or AhR antagonist (10 μM, EDM Millipore), were added to the culturing media. All cytokines were purchased from R&D. Click's (Irvine Scientific) or RPMI (SIGMA-ALDRICH) (when indicated) media were supplemented with 10% FBS, L-glutamine (2 mM), penicillin (100 U ml^{-1}) and β-mercaptoethanol (40 nM). After 4–5 days of culture, the cells were acquired at the FACS.

Foxp3^{RFP} IL-17A^{Katushka} IL-10^{eGFP} triple reporter were injected with anti-CD3 mAb and 12 h after the first injection or 4 h after the third injection a pure population of intestinal CD4$^+$ Foxp3^{RFP-} IL-17A$^{Katushka+}$ IL-10^{eGFP-} cells were FACS sorted and restimulated *in vitro* in the presence of irradiated splenocytes (1:4 ratio). The cells were stimulated for 5 days in the presence of soluble anti-CD3 monoclonal antibody (2 μg ml^{-1}), IL-6 (20 ng ml^{-1}) and where indicated anti-TGF-β (5 μg ml^{-1}), TGF-β (0.25 ng) and FICZ (100 nM).

RNA amplification, extraction and sequencing. We isolated intestinal lymphocytes from two independent experiments, each using 5 mice injected with anti-CD3 monoclonal antibody. The cell populations indicated in Fig. 2 were FACS-sorted from these two independent experiments and the cells of each population were pooled before the RNA extraction, amplification and sequencing. Around 5,000 cells for each population were processed. After sorting, the cells were washed 2 times (1,500 r.p.m., 2 min, 4 °C) with 1 ml phosphate-buffered saline (PBS) and finally suspended in 2.5 μl PBS (containing 0.5 μl RNaseOut (Invitrogen) and 0.5μl dithiothreitol (DDT) (Invitrogen)). After that, the cytoplasm RNA was isolated as described previously[33,34]. Briefly, 2.5 μl of 2x selected cytoplasm lysis buffer (SCLB) was added and the cells were lysed by pipetting up and down for 5 times. The entire lysate solution was spun at 8,000 r.p.m. for 5 min in a chilled centrifuge (4 °C) and the supernatant (~5 μl), which contained the total cytoplasm RNA, was transferred to a PCR tube strip with individually attached dome caps (USA Scientific). The mRNA selection, reverse transcription and cDNA amplification was performed as described previously with some modifications[33,34]. Briefly, 5′-phosphorylated oligo-GdT24 (pGdT24) primer was used to selectively reverse transcribe mRNAs. The first strand cDNA was synthesized with Superscript Reverse Transcriptase III (Invitrogen). Then, the double-stranded cDNA (dscDNA) was generated. The cDNA was purified with the Genomic DNA Clean & Concentrator kit (Zymo). Afterwards, several steps including DNA end-blunting, 5′-end phosphorylation and ligation were performed with The End-It DNA End-Repair Kit (Epicentre) with T4 DNA ligase (Epicentre). The product was directly amplified using REPLI-g UltraFast Mini Kit (QIAGEN) and purified using the same Genomic DNA Clean & Concentrator kit (Zymo) column. Finally, 3 to 5 μg (up to 8 μg) of amplified cDNA derived from mature mRNA was obtained, which was then evaluated by PCR and fragmented to construct sequencing library.

We follow the standard Illumina HiSeq2000 protocol to make the library. Briefly, the amplicon was fragmented to an approximately 200–500 bp size range by a Bioruptor Sonicator (Diagenode). After purification with DNA Clean & Concentrator kit (Zymo), end-repairing, 3′-A tailing and ligation with adaptor was performed. Then, a 50 bp range DNA (250–300 bp) was selected by gel electrophoresis (E-gel EX 2%, Invitrogen) and barcode added by PCR using Phusion High-Fidelity DNA polymerase (NEB) for 8 cycles. The product was size selected again and the DNA concentration was quantitated by a Bioanalyzer (Agilent). Multiple samples were mixed and loaded to the Hi-Seq2000 for sequencing performed with 50 bp single-end reads.

RNA-Seq data were aligned to the *Mus musculus* GRCm37 genome using Tophat2 and default settings[35]. Duplicate reads were removed with samtools rmdup command[36]. Count data was generated using HTSeq-count and FPKM data was generated with Cufflinks[37,38]. To determine genes important for the T_R1 exT_H17 conversion we performed a differential expression test using DESeq2 comparing T_H17 cells with T_R1 exT_H17 and T_R1 cells. To determine which cell populations were more closely related, pearson correlation values between samples were calculated on $\log_2$ FPKM data after a maximum filter of FPKM >1, based on the subset of genes listed in Supplementary Table 1 and 2. In order to determine the grouping of cell populations, hierarchical cluster analysis was performed on the distances based on correlation or gene expression. The linkage criteria complete was used for clustering analysis.

Real time PCR. Total RNA was extracted from cells using TRIzol reagent. To synthetize the cDNA we then used the High Capacity cDNA Sythesis Kit (Applied Biosystem) and RT–PCR was performed using the TaqMan Fast Universal PCR Master Mix and TaqMan Gene Probes (Applied Biosystems) on a 7500 Fast Real-time PCR system machine (Applied Biosystem). Samples were run in duplicate or triplicate and expression levels were calculated as relative to the expression of endogenous *HPRT* or *Polr2a*.

Experimental autoimmune encephalomyelitis and tetramers staining. Mice were immunized sub-cutaneously with an emulsion of 250 mg of $MOG_{35\text{-}55}$ peptide (Yale Keck facility) and CFA (BD Difco). At the time of immunization and 48 h after, mice received 200 ng pertussis toxin (PTx, List Biological Laboratories) per each injection. The clinical score of EAE development was addressed daily according to guidelines: 0, no signs of disease; 0,5, tail weakness; 1, complete tail paralysis; 2, partial hind limb paralysis; 2,5, unilateral complete hind limb paralysis; 3, complete bilateral hind limb paralysis; 3,5, complete hind limb paralysis and partial forelimb paralysis; 4, total paralysis of forelimbs and hind limbs, moribund. All mice experiments were conduced according to IACUC policies. To identify $MOG_{38\text{-}49}$ (mouse myelin oligodendrocyte glycoprotein 38-49, "GWYRSPFSRWH") specific T-cells, 10^7 cells per ml were incubated with neauraminidase (0.5 U ml^{-1}, neuraminidase type X from *Clostridium perfringens*, Sigma) in serum-free DMEM at 37 °C/5% CO_2 for 25min. After this, the cells were stained with the $MOG_{38\text{-}49}$/I-A(b)-tetramer allophycocyanin (APC)-labelled (NIH Tetramer Facility) for 4 h, at room temperature in DMEM, 2%FBS. Cells were then stained for surface antigens and acquired at the FACS.

***Nippostrongylus brasiliensis* infection and isolation of lymphocytes from the lung.** Third-stage larvae (L3) of *N. brasiliensis* were recovered from cocultures of infected mice. We infected mice by injecting subcutaneously 625 parasites in 0.2 ml PBS at the base of the tail, as previously described[9]. Mice were euthanized at different time points, as described, and lymphocytes were isolated from lungs. Cell suspensions from lungs were obtained digesting the organs, previously cut in small pieces, in 10% FBS RPMI media in the presence of DNase and collagenase D, as previously described[9]. After digestion, cell suspensions were processed onto a Percoll gradient (40% on 100%) and lymphocytes were then processed for FACS analysis.

***Staphylococcus aureus* infections.** *S. aureus* (ATCC 14458, SEB^+ $TSST\text{-}1^-$) was injected intravenously (10^{10} colony-forming units per mouse) into $Fate^+$ mice not older than 7 weeks. Mice were killed 3–4 days after the injection, at a time when they displayed severe clinical symptoms of sepsis (weight loss, dehydration, lethargy) and the presence of T_H17, exT_H17 and $T_R1^{exT_H17}$ cells was tested in the small intestine by FACS.

T_H17 transfer colitis, endoscopic and histologic analysis. $CD4^+$ were isolated from the IL-17A^{eGFP} Foxp3^{RFP} double report mice and cultured with irradiated antigen presenting cells (1:4 ratio), in the presence of soluble anti-CD3 mAb (2C11, 1 µg ml^{-1}), IL-6 (20 ng ml^{-1}), IL-23 (50 ng ml^{-1}) and TGF-β1 (0.25 ng ml^{-1}) along with neutralizing antibodies for IFN-γ (XMG1.2 clone, 10 µg ml^{-1}) and IL-4 (11B11 clone, 10 µg ml^{-1}). After 5 days of *in vitro* culture, a pure population of 10,000 pathogenic (p) T_H17 cells (FACS sorted as $CD4^+$ IL-17^{eGFP+} Foxp3^{RFP-}) were injected intra peritoneally into *Rag1*$^{-/-}$ mice at 1:1 ratio with the following cell populations: T_H17, exT_H17, T_R1 exT_H17 and T_R1 cells. These later populations were FACS sorted from the intestine of $Fate^+$ 4 h after the second injections of anti-CD3 monoclonal antibody.

When indicated (Fig. 4c, d) both T_H17 and $T_R1^{exT_H17}$ cells were generated *in vitro* under the same condition (CD3 (10 µg ml^{-1}) and CD28 (1–2 µg ml^{-1}); TGF-β1 (0.25–0.5 ng ml^{-1}), IL-6 (20 ng ml^{-1}), IL-23 (20 ng ml^{-1}), and antibodies to IFN-γ (10 µg ml^{-1}) and IL-4 (10 µg ml^{-1})) and then sorted and transferred ($n = 10,000$ to 20,000 cells) into *Rag1*$^{-/-}$ mice.

When indicated (Extended Data Fig. 9), the $T_R1^{exT_H17}$ cells were generated *in vitro* in the presence of TGF-β1 (1 ng ml^{-1}), IL-6 (20 ng ml^{-1}), IL-23 (20 ng ml^{-1}), and antibodies to IFN-γ (XMG1.2, 10 µg ml^{-1}) and IL-4 (11B11, 10 µg ml^{-1}) and FICZ (100 nM; Enzo Life Sciences), FACS sorted and transferred ($n = 10,000$ cells) into *Rag1*$^{-/-}$ mice at 1:1 ratio with pT_H17 cells.

Colonoscopy was performed in a blinded fashion using the Coloview system (Karl Storz, Germany). Briefly, colitis score was addressed considering the consistence of stools, granularity of the mucosal surface, translucency of the colon, fibrin deposit and vascularization of the mucosa (0–3 points for each parameter). Haematoxilin and eosin staining were performed on paraffin sections of colon previously fixed in Bouin's fixative solutions.

Statistical analysis and FACS analysis. Statistical analysis were performed using Prism 5.0 (Graphpad Software) Paired *t* test, Non parametric Mann–Whitney *U*-test, ANOVA (post test Tukey, Bonferroni or Dunnet) were used according to the type of experiments. $\log_{10}$-transformed values for cell counts were used in Fig. 1. P-values ≤ 0.05 were considered significant (*: $P < 0.05$; **: $P < 0.005$; ***: $P < 0.0005$); P values >0.05; non-significant (NS). All flow cytometry data have been analysed with FlowJo (Treestar). No statistical methods were used to predetermine sample size.

31. Maynard, C. L. *et al.* Regulatory T cells expressing interleukin 10 develop from Foxp3$^+$ and Foxp3$^-$ precursor cells in the absence of interleukin 10. *Nature Immunol.* **8,** 931–941 (2007).
32. Jinnin, M., Ihn, H. & Tamaki, K. Characterization of SIS3, a novel specific inhibitor of Smad3, and its effect on transforming growth factor-beta1-induced extracellular matrix expression. *Mol. Pharmacol.* **69,** 597–607 (2006).
33. Guo, S. *et al.* Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156,** 649–662 (2014).
34. Pan, X. *et al.* Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc. Natl Acad. Sci. USA* **110,** 594–599 (2013).
35. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
36. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
37. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
38. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2014).
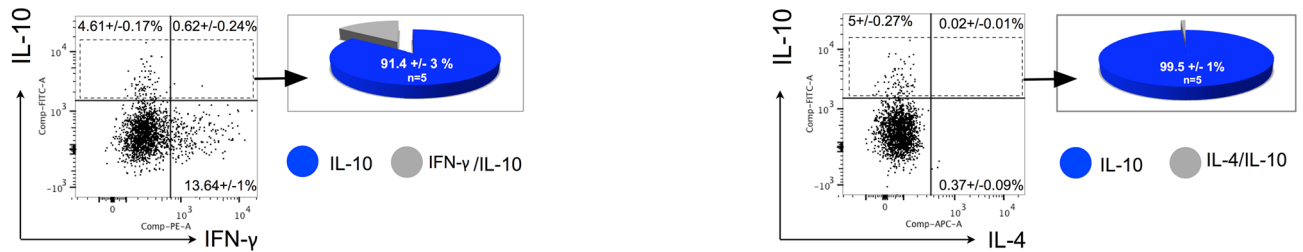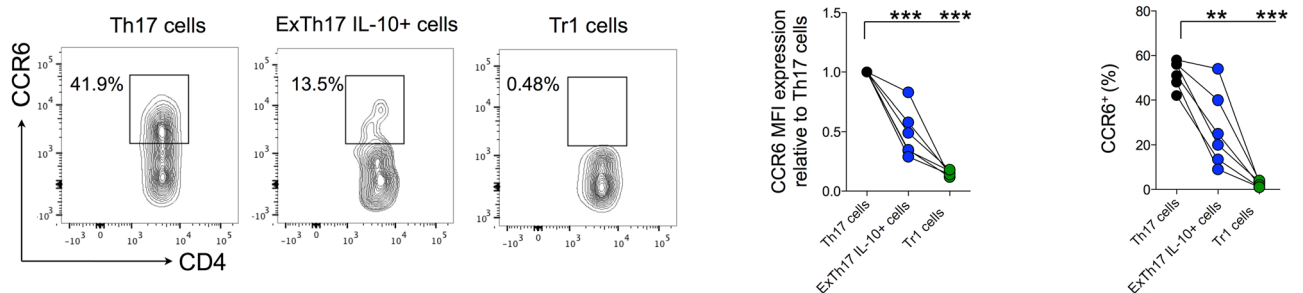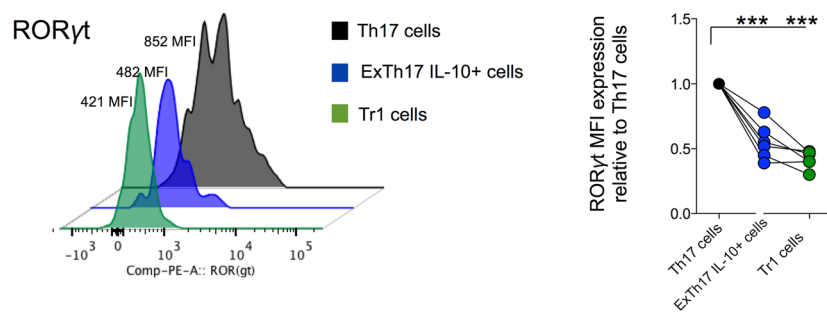
a

Fate+ mouse model
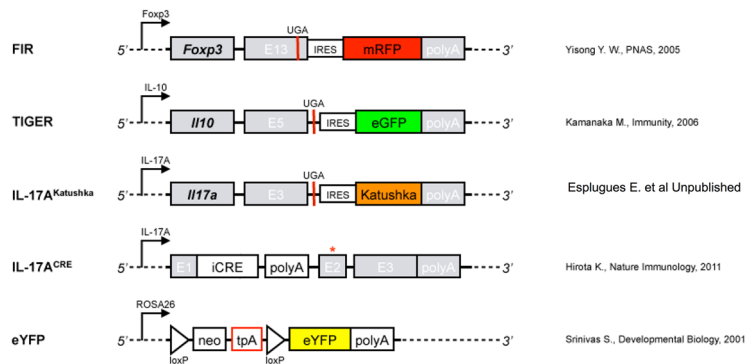

b
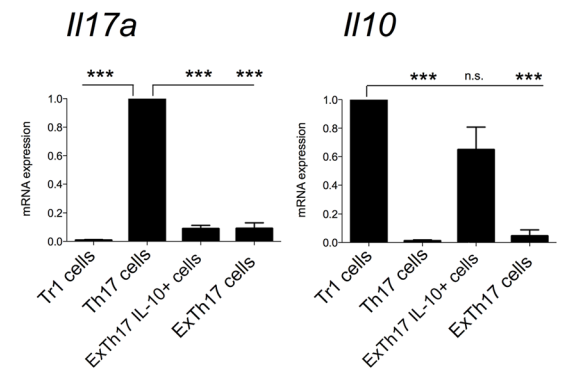
c

d

e

Extended Data Figure 1 | Description of Fate$^+$ mice and characterization of exT$_H$17 IL-10^{eGFP+} cells under steady state condition. a, Constructs contained in the Fate$^+$ mice. **b**, During anti-CD3 mAb induced transient inflammation in the S.I., a sufficient number of exT$_H$17 IL-10^{eGFP+} was generated to test whether Fate$^+$ mice faithfully report IL17A and IL-10 expression. In particular, T$_R$1 (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^-$ IL-10^{eGFP+} Foxp3^{RFP-}), T$_H$17 (CD4$^+$ IL-17A$^{Katushka+}$ YFP$^+$ IL-10^{eGFP-} Foxp3^{RFP-}), exT$_H$17 IL-10^{eGFP+} (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^+$ IL-10^{eGFP+} Foxp3^{RFP-}) and exT$_H$17 (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^+$ IL-10^{eGFP-} Foxp3^{RFP-}) were FACS sorted from the small intestine of anti-CD3 monoclonal antibody treated Fate$^+$ mice and mRNA expression relative to T$_H$17 cells for *Il17a* and relative to T$_R$1 for *Il10* is reported. ExT$_H$17 IL-10^{eGFP+} express *Il10*$^{dim/high}$ and *Il17a*^{low}. Data are cumulative of three independent experiments. In each experiment we pooled cells from at least 7 mice. Mean and s.e.m., ***$P \le 0.0005$ by ANOVA (Tukey's multiple comparison test). **c–e**, Under steady state conditions, intestinal lymphocytes were isolated and re-stimulated *in vitro* for 3 h with PMA/ionomycin for the intracellular staining of IFN-γ and IL-4, while they were freshly analysed for the expression of CCR6 and RORγt. **c**, Frequencies of IFN-γ and IL-4 among the exT$_H$17 cells is shown. Pie chart reports the frequencies of the indicated cytokine among the exT$_H$17 IL-10^{eGFP+}. One biological replicates out of five is shown. **d**, **e**, Frequencies and MFI of CCR6 (**d**) and MFI of RORγt are reported for T$_R$1 (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^-$ IL-10^{eGFP+}), T$_H$17 (CD4$^+$ IL-17A$^{Katushka+}$ YFP$^{+/-}$ IL-10^{eGFP-}), exT$_H$17 IL-10$^+$ (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^+$ IL-10^{eGFP+}) (**e**). Each dot represents one biological replicate. Mean and s.e.m., **$P \le 0.005$, ***$P \le 0.0005$ by ANOVA (Dunnett's multiple comparison test, comparison all columns vs control (T$_H$17 cells)).
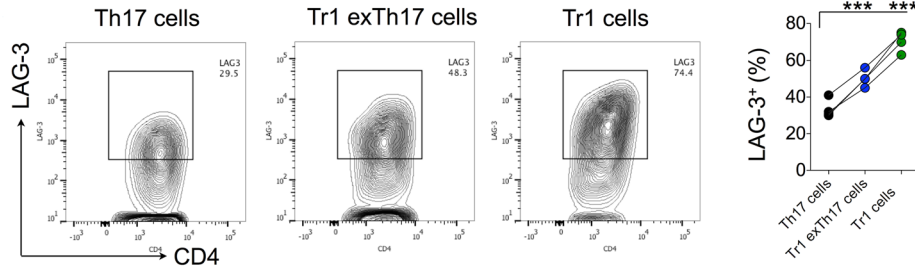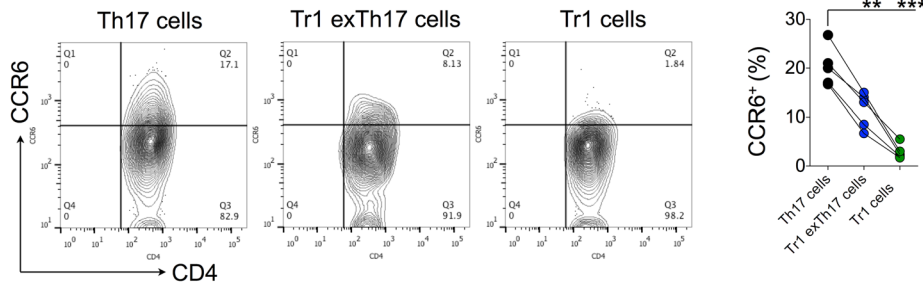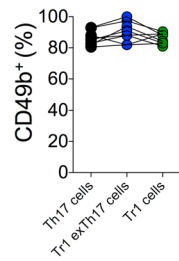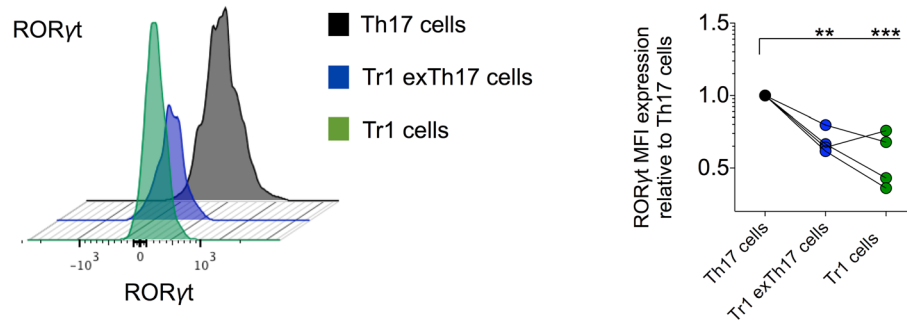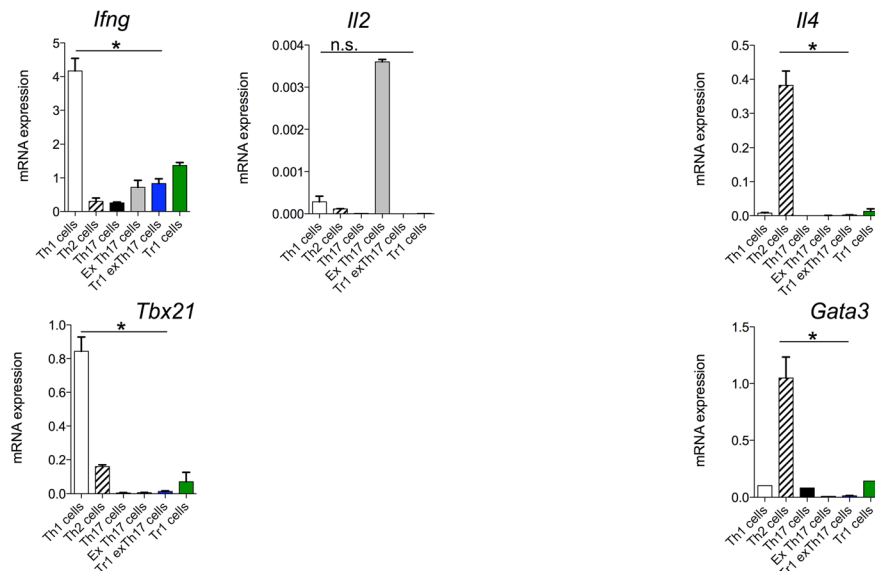
Extended Data Figure 2 | Characterization of Tr1^exTH17 cells. Tr1 (CD4+ IL-17A^Katushka− YFP− IL-10^eGFP+), TH17 (CD4+ IL-17A^Katushka+ YFP^+/− IL-10^eGFP−), Tr1^exTH17 (CD4+ IL-17A^Katushka− YFP+ IL-10^eGFP+) were isolated from the small intestine of anti-CD3 monoclonal antibody treated-Fate+ mice and analysed by FACS. **a–d,** Frequencies of LAG-3 (**a**), CCR6 (**b**), CD49b (**c**) and MFI of RORγt (**d**) are reported. Each dot represents one biological replicate. Mean and s.e.m., **P ≤ 0.005, ***P ≤ 0.0005 by ANOVA (Dunnett's multiple comparison test, Comparison all columns vs control (TH17 cells)).

e, TH1 (CD4+ IFN-γ^Katushka+ Foxp3^RFP−), TH2 (CD4+ IL-4^GFP+ Foxp3^RFP−), Tr1 (CD4+ IL-17A^Katushka− IL-10^eGFP+ Foxp3^RFP−), TH17 (CD4+ IL-17A^Katushka+ Foxp3^RFP−), Tr1^exTH17 (CD4+ IL-17A^Katushka− YFP+ IL-10^eGFP+ Foxp3^RFP−) and exTH17 (CD4+ IL-17A^Katushka− YFP+ IL-10^eGFP− Foxp3^RFP−) were FACS sorted from the small intestine of anti-CD3 monoclonal antibody treated mice. mRNA expression relative to *HPRT* of *Ifng, Il2, Tbx21, Il4* and *Gata3* of the indicated populations is reported. Mean and s.e.m., *P ≤ 0.05, by Mann–Whitney U-test, two tailed.
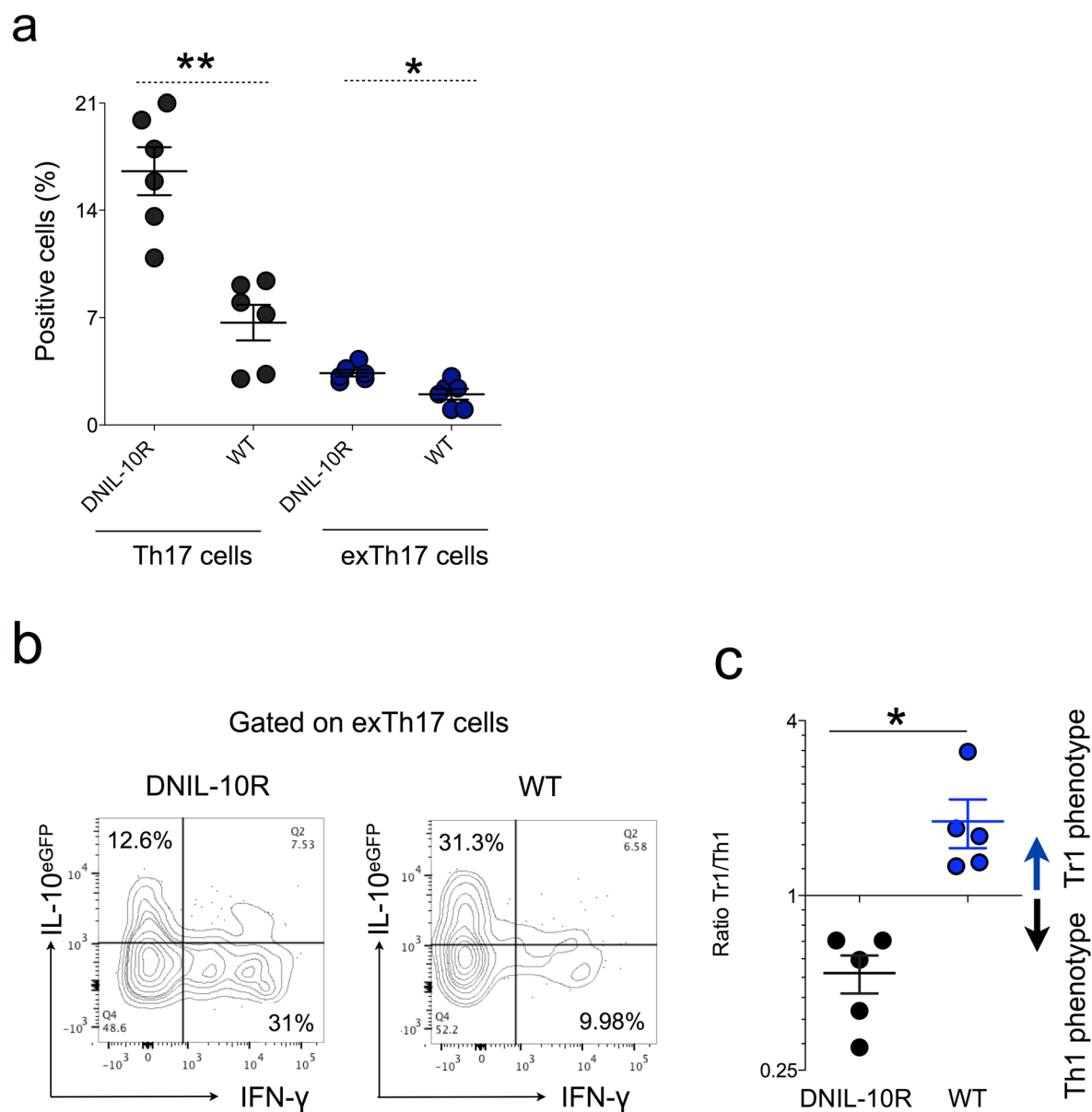
Extended Data Figure 3 | TH17 fate in DNIL-10R Fate mice. Fate (WT) and Fate dominant negative IL-10R (DNIL-10R) were injected with anti-CD3 monoclonal antibody and the fate of small intestinal TH17 cells analysed. **a**, Frequencies of TH17 and exTH17 gated on CD4$^+$ T cells are reported. **b**, Representative flow cytometric analysis of the IL-10 and IFN-γ expression in exTH17 cells. **c**, Ratio between TR1^{exTH17} (CD4$^+$ IL-17A$^-$ YFP$^+$ IL-10^{eGFP+} IFN-γ$^-$) and TH1^{exTH17} (CD4$^+$ IL-17A$^-$ YFP$^+$ IL-10^{eGFP-} IFN-γ$^+$) in WT and DNIL-10R mice. Mean ± s.e.m. Each dot represents one biological replicate. *$P \leq 0.05$; **$P \leq 0.005$ by Mann–Whitney U-test, two tailed.
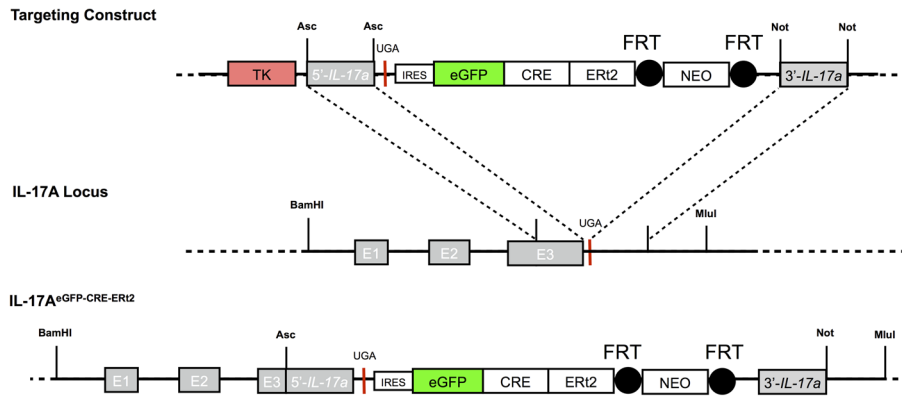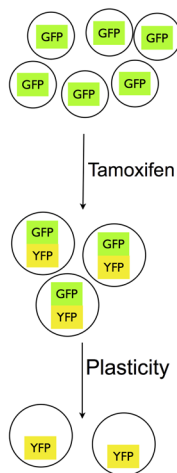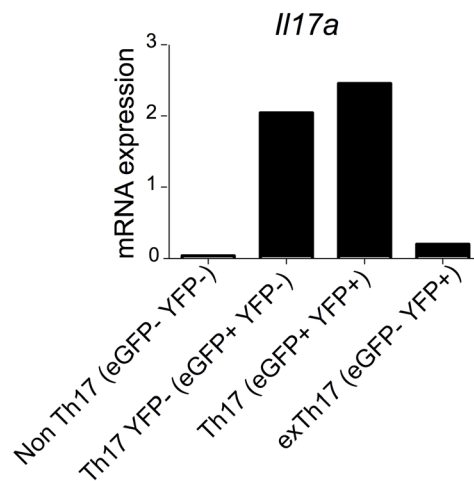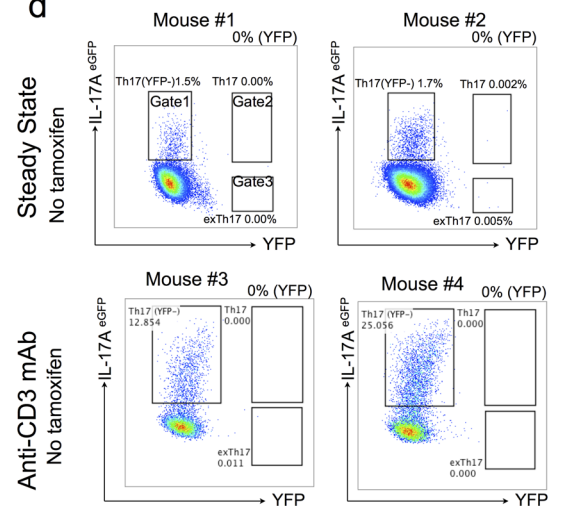
Extended Data Figure 4 | Constructs and validation of iFate mice.
a, Targeting strategy and constructs of IL-17A^eGFP-CRE-ERT2 mice (iFate). Black circles represent Flp recombinase target (FRT) sites. **b**, Schematic of TH17 cell development in iFate showing that TH17 cell plasticity can be tested only after tamoxifen treatment. **c**, iFate mice were injected with tamoxifen and anti-CD3 monoclonal antibody and non TH17 cells (IL-17A^GFP−YFP−), TH17 YFP^+ cells (IL-17A^eGFP+ YFP−), TH17 cells (IL-17A^eGFP+ YFP^+) and exTH17 cells (IL-17A^eGFP− YFP^+) were FACS sorted. *Il17a* mRNA expression in the indicated populations is reported. The mRNA expression is normalized to *HPRT*. One representative experiment out of two is shown. **d**, Representative flow cytometric analysis of intestinal TH17 cells under steady state condition or after anti-CD3 monoclonal antibody (aCD3) in the absence of tamoxifen treatment. The efficiency of CRE-mediated recombination after tamoxifen is reported as frequencies of YFP^+ cells (YFP) and is the result of the following calculation: YFP^+ cells (gate 2 + 3) / (IL-17A^eGFP+ YFP− cells (gate 1) + YFP^+ cells (gate 2 + 3).
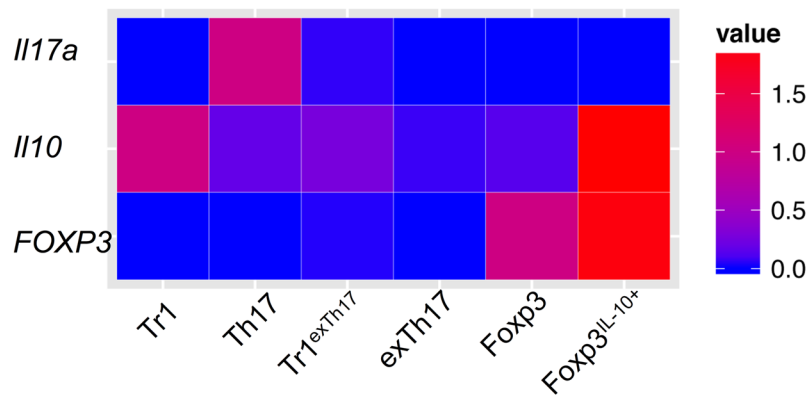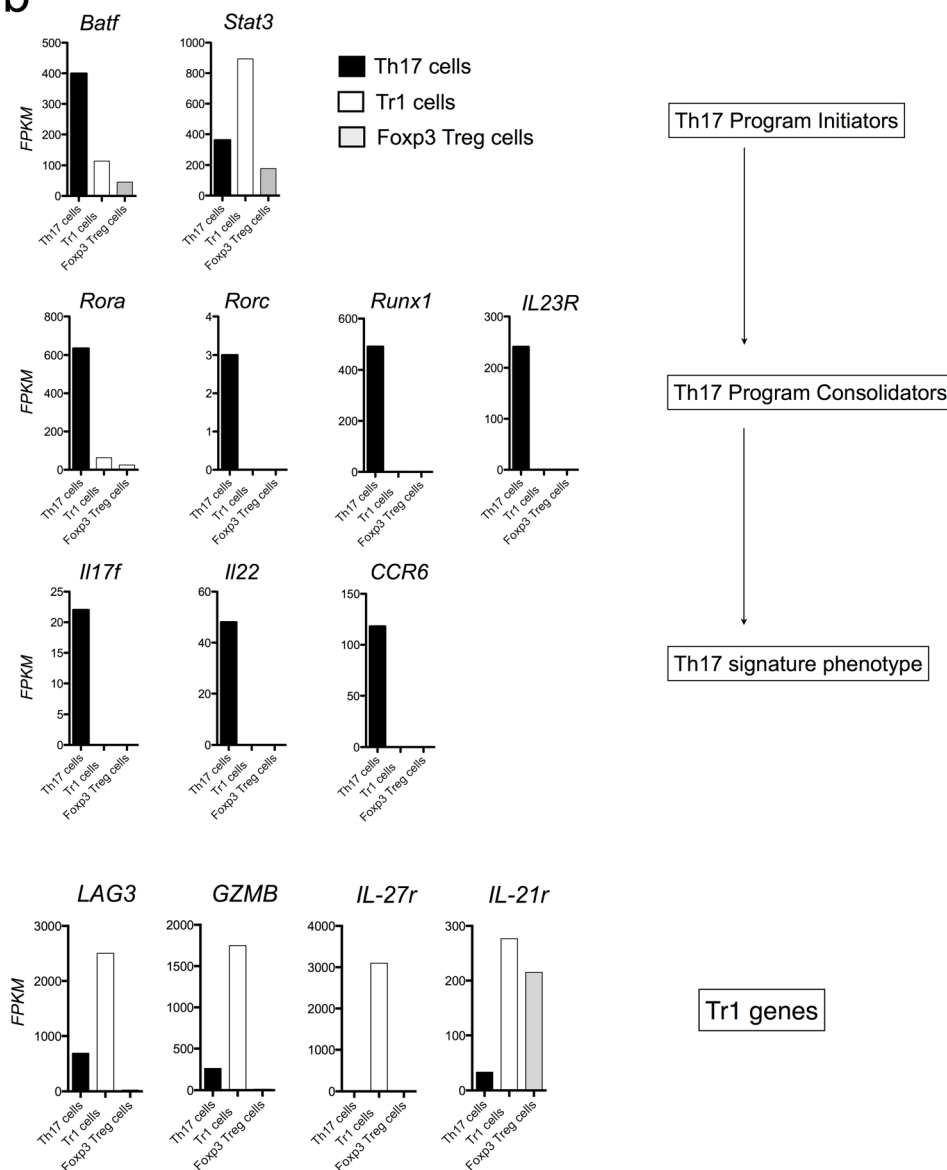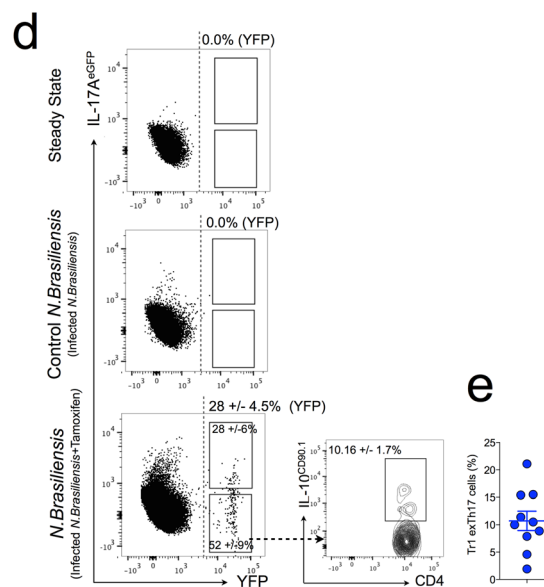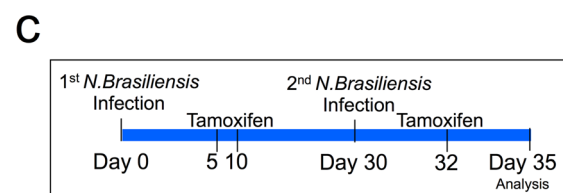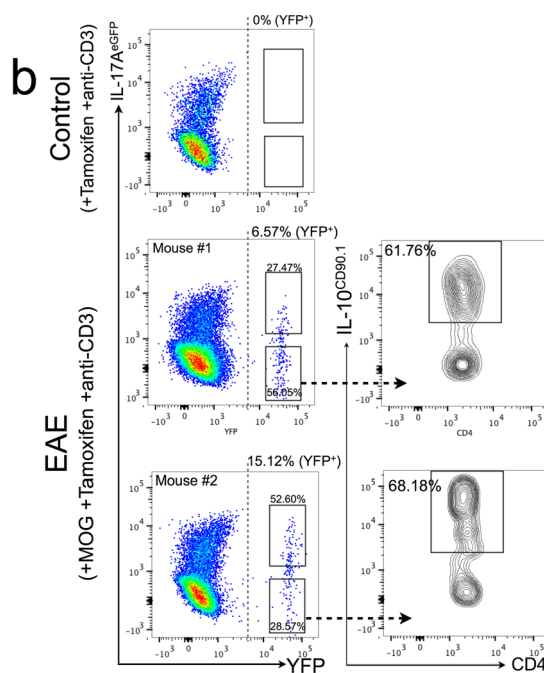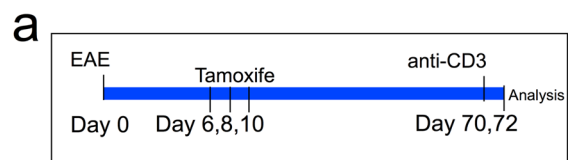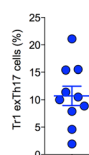
Extended Data Figure 5 | Relative expression of *Il17a*, *Il10* and *Foxp3* and FPKM values of signature genes of bona fide Th17 and Tr1 cells. a, Tr1 (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^-$ IL-10^{eGFP+} Foxp3^{RFP-}) Th17 (CD4$^+$ IL-17A$^{Katushka+}$ YFP$^{+/-}$ IL-10$^{eGFP+/-}$ Foxp3^{RFP-}), Tr1^{exTh17} (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^+$ IL-10^{eGFP+} Foxp3^{RFP-}), exTh17 (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^+$ IL-10^{eGFP-} Foxp3^{RFP-}) Foxp3$^+$ T$_{Reg}$ (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^-$ IL-10^{eGFP-} Foxp3^{RFP+}) and Foxp3$^+$ T$_{Reg}$ IL-10$^+$ (CD4$^+$ IL-17A$^{Katushka-}$ YFP$^-$ IL-10^{eGFP+} Foxp3^{RFP+}) cells were isolated from the small intestine of Fate$^+$ mice after anti-CD3 monoclonal antibody injections. The transcriptome of these populations was sequenced and the relative FPKM expressions of *Il10*, *Il17a* and *Foxp3* compared to Tr1, Th17 and Foxp3$^+$ T$_{Reg}$ cells are reported. **b**, FPKM values of the indicated populations of the reported genes are shown.

a

EAE   Tamoxife   anti-CD3

Day 0   Day 6,8,10   Day 70,72   Analysis

b

Control (+Tamoxifen +anti-CD3)

IL-17A^eGFP

0% (YFP⁺)

EAE (+MOG +Tamoxifen +anti-CD3)

Mouse #1   6.57% (YFP⁺)
27.47%
56.05%

IL-10^CD90.1   61.76%
CD4

Mouse #2   15.12% (YFP⁺)
52.60%
28.57%

68.18%
CD4

YFP

c

1st *N.Brasiliensis* Infection   2nd *N.Brasiliensis* Infection

Tamoxifen   Tamoxifen

Day 0   5 10   Day 30   32   Day 35
Analysis

d

Steady State

IL-17A^eGFP

0.0% (YFP)

Control *N.Brasiliensis* (Infected *N.Brasiliensis*)

0.0% (YFP)

N.Brasiliensis (Infected *N.Brasiliensis*+Tamoxifen)

28 +/- 4.5% (YFP)
28 +/-6%
52 +/-9%

YFP

IL-10^CD90.1   10.16 +/- 1.7%
CD4

e

Tr1 exTh17 cells (%)

25
20
15
10
5
0

Extended Data Figure 6 | Tr1^exT_H17 cell development in EAE and during helminth infection using iFate mice. **a**, Schematic of the experiment, showing iFate^+ mice immunized with MOG, treated for 3 times with tamoxifen and then injected with anti-CD3 monoclonal antibody 70 and 72 days after MOG immunization. The intestinal lymphocytes were analysed 4 h after the second injection of anti-CD3 monclonal antibody. **b**, Representative flow cytometric analysis of T_H17 and exT_H17 (gated on YFP^+ cells) and Tr1^exT_H17 cells (gated on exT_H17). The YFP^+ percentages (YFP) shown on the dot plots report the efficiency of tamoxifen-induced CRE-recombination. Three representative biological replicates out of six are shown. **c**, Schematic of the experiment, showing iFate mice infected with *N. brasiliensis* and injected i.p. with tamoxifen at the indicated time points. **d**, Representative flow cytometric analysis of CD4^+

T cells isolated from the lung of iFate mice before (steady state) and after the second infection ± tamoxifen (control *N. brasiliensis* (no tamoxifen) and *N. brasiliensis* (+ tamoxifen) respectively)). Cumulative dot plots of 3 biological replicates are shown. One representative experiments out of 3 is shown. The YFP^+ percentages (YFP) shown on the dot plots report the efficiency of tamoxifen-induced CRE-recombination. The frequencies within the cumulative dot/density plot report the percentage of T_H17 and exT_H17 among the YFP^+ cells, and the frequency of IL-10^+ cells among the exT_H17 cells. **e**, Frequencies of Tr1^exT_H17 cells (gated on exT_H17). Each dot represents a biological replicates. Results are cumulative from three independent experiments.
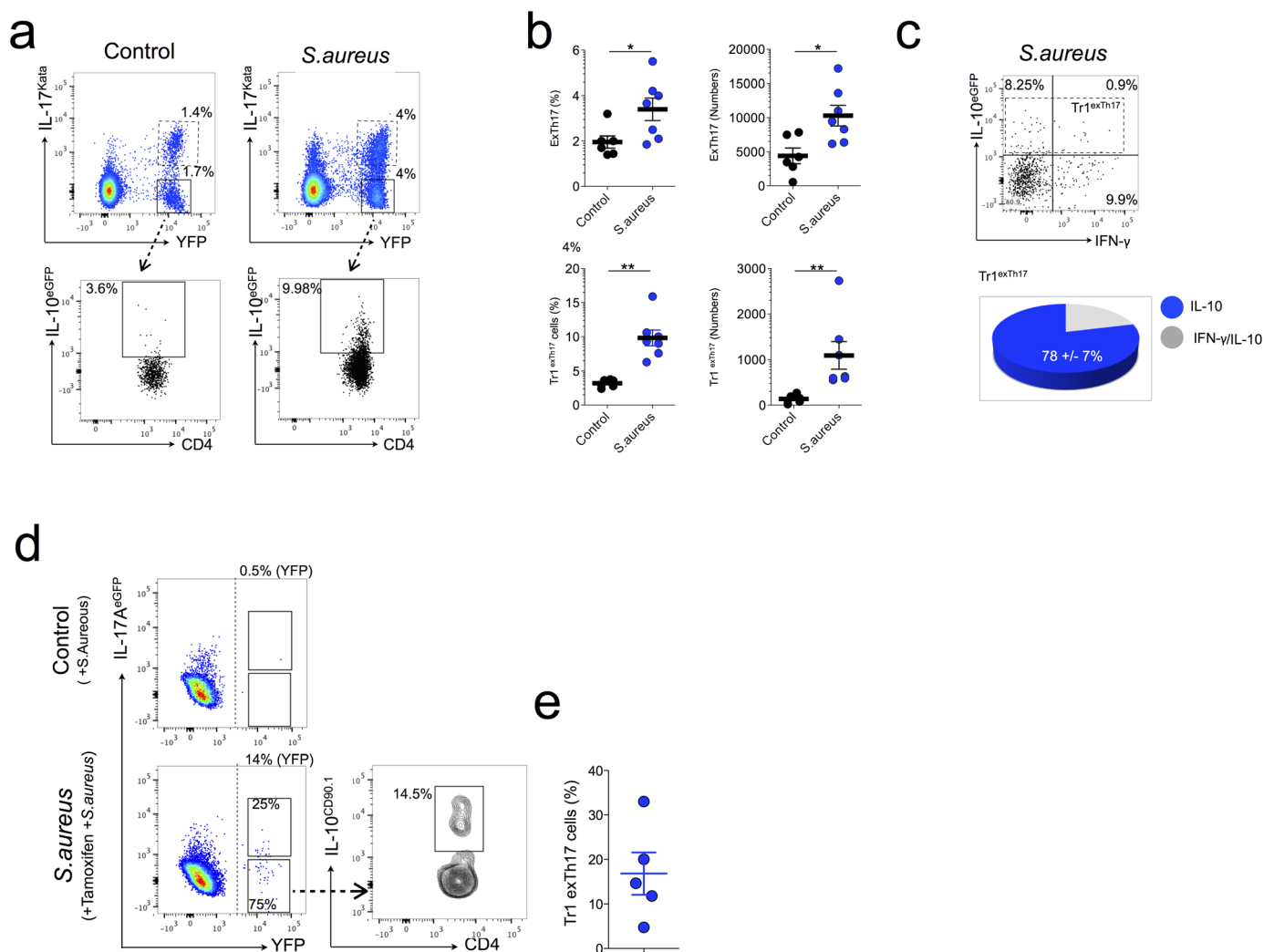
Extended Data Figure 7 | Conversion of TH17 cells into TR1 over the course of *S. aureus* infection using Fate and iFate mice. **a**, Fate[+] mice were left untreated (control) or injected i.v. with *S. aureus* (*S. aureus*). Representative flow cytometric analysis of intestinal TH17 and exTH17 (gated on CD4[+] Foxp3[RFP−]) and TR1[exTH17] cells (CD4[+] IL-17A[Katushka−] YFP[+] IL-10[eGFP+]; gated on exTH17) are shown. One representative experiment out of three is shown. **b**, Frequencies and numbers of the indicated population in the small intestine of untreated (control) and infected mice (*S. aureus*) are reported. Results are cumulative from three independent experiments. Mean and s.e.m., **$P \leq$ 0.005, ***$P \leq$ 0.0005 by Mann–Whitney U-test, two tailed. **c**, IFN-γ and IL-10[eGFP] expression of exTH17 cells. Pie chart reports the frequencies of the indicated cytokine among the TR1[exTH17] cells. **a–c**, One representative biological replicate out of three is shown. One representative experiment out of two is shown. **d**, Representative flow cytometric analysis of intestinal lymphocytes isolated from iFate mice 4 days after *S. aureus* infection. One representative biological replicate out of 5 is shown. The YFP[+] percentages (YFP) shown on the dot plots report the efficiency of tamoxifen-induced CRE-recombination. **e**, Frequencies of TR1[exTH17] cells (CD4[+] IL-17A[eGFP−] YFP[+] IL-10[CD90.1+]; gated on exTH17) are shown. Results are cumulative from two independent experiments.

a

Color Key

0   4   8
Value

b
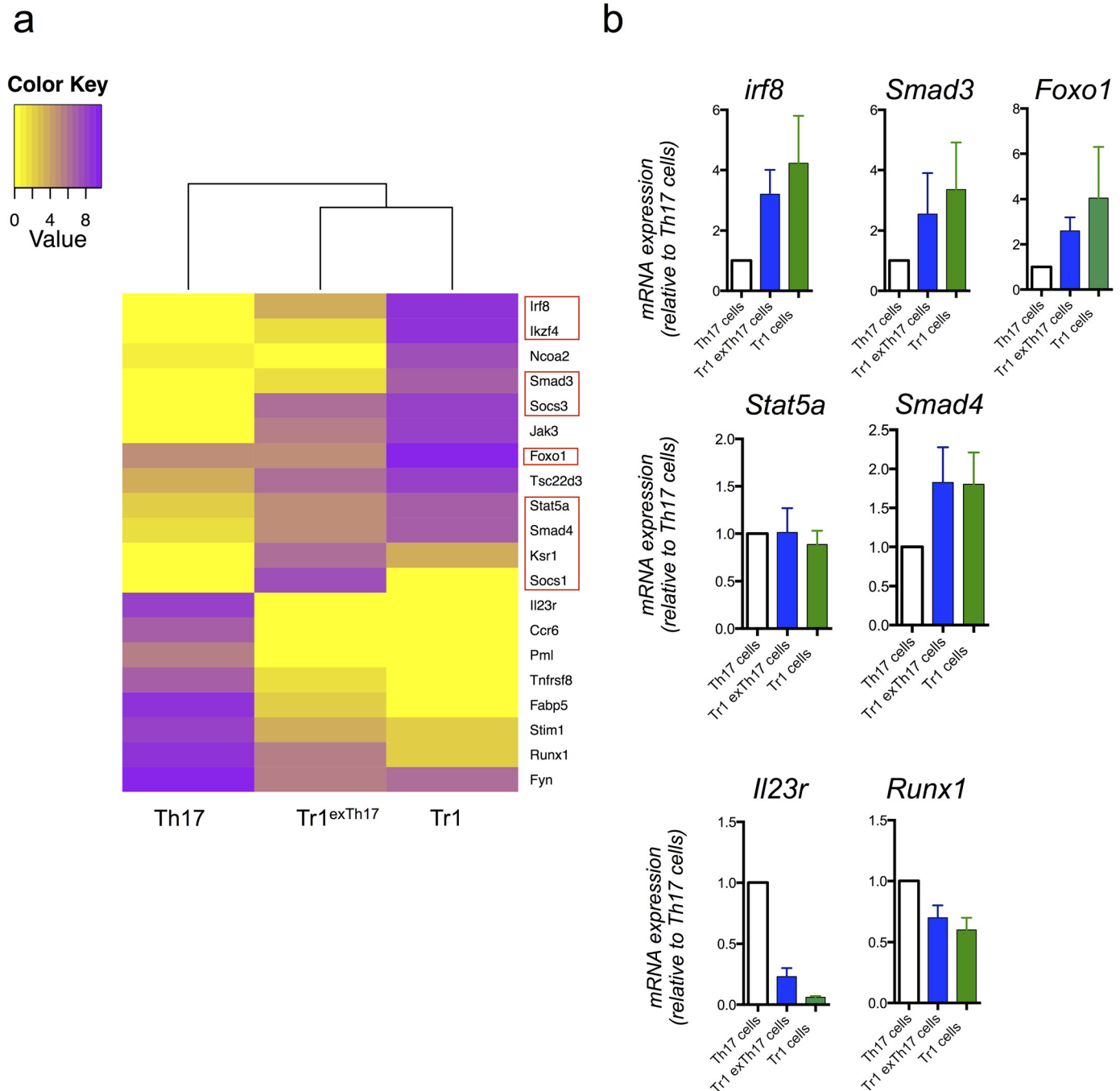
Extended Data Figure 8 | Gene expression of TR1, TR1^exTH17 and TH17 cells.
a, Heat map of genes selectively expressed in both TR1^exTH17 and TR1 compared to TH17 cells. The bioinformatics analysis is based on the genes listed in Supplementary Table 1. Red squares highlight genes linked to TGF-β1 signalling. **b**, Relative mRNA expression of the indicated genes in TR1, TR1^exTH17 and TH17 cells FACS sorted form the intestine of Fate+ mice treated with anti-CD3 monoclonal antibody is shown. Values shown are relative to TH17 cell gene expression. Mean and s.e.m. of biological independent experiments (IRF8 $n = 2$; SMAD3 $n = 4$; FOXO1 $n = 2$; STAT5a $n = 3$; SMAD4 $n = 4$) except for IL-23 and Runx1 ($n = 2$ technical replicates) are shown. In each experiment we pooled intestinal lymphocytes isolated from 7 treated mice before FACS sorting.
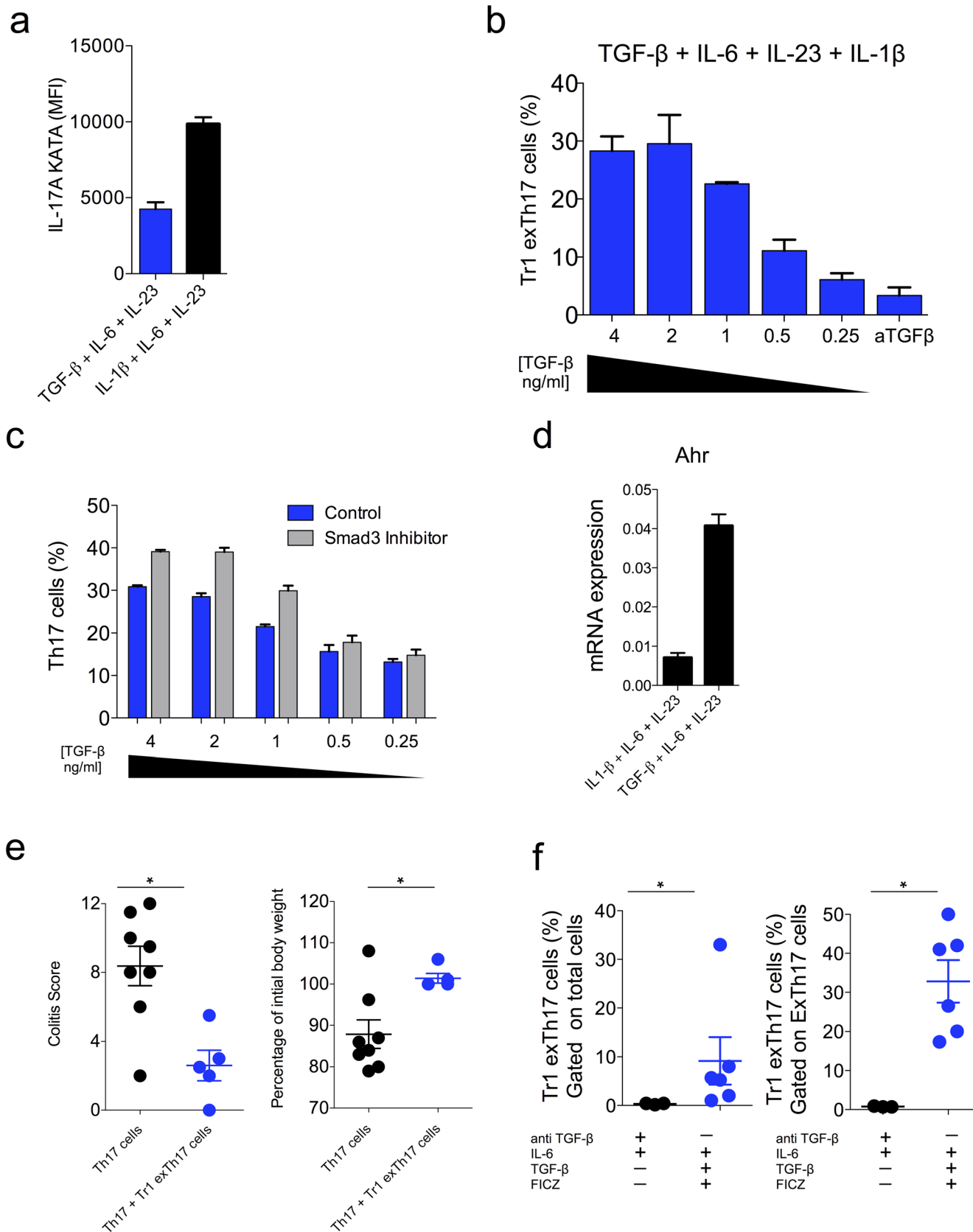
Extended Data Figure 9 | Characterization of *in vitro* generated TR1^{exTH17} cells. a, IL-1β counteracted TH17 plasticity. IL-17A MFI in TH17 (CD4^{+} Foxp3^{RFP-} IL-17A$^{Katushka+}$ YFP$^{+/-}$ IL-10$^{eGFP+/-}$) differentiated in the presence of TGF-β1, IL-6, IL-23 or IL-1β, IL-6, IL-23. One experiment out of five is shown. Two technical replicates are reported. **b**, Dose-response effect of TGF-β1 on the induction of TR1^{exTH17} cells cultured in the presence of IL-6, IL-23, IL-1β. In the last conditions we added anti-TGF-β1 monoclonal antibody. TGF-β1 was diluted 1:2 starting from the concentration of 4 ng ml^{-1}. One experiment out of five is shown. Two technical replicates are reported. **c**, In line with the literature[29], Smad3 chemical inhibition also favours TH17 cell development. Frequency of TH17 cells cultured in the presence or in the absence Smad3 inhibitor at the indicated different concentrations of TGF-β1 (4–0.25 ng ml^{-1}). One experiment out of five is shown. Three technical replicates are reported. **d**, mRNA expression of *Ahr* in CD4 T cells cultured in the presence of either TGF-β1, IL-6, IL-23 or IL-1β, IL-6, IL-23. The expression is normalized to HPRT. One experiment out of two is shown. Two technical replicates are reported. **e**, TR1^{exTH17} cells were polarized *in vitro* in the presence of TGF-β1+IL-6+IL-23+FICZ and transferred into *Rag1*$^{-/-}$ mice ± (p)TH17 cells. Endoscopic colitis score and percentage of initial body weight in the indicated groups are shown. Each dot represents one mouse. Results are cumulative from three independent experiments. Mean and s.e.m., *$P \le 0.05$ by Mann–Whitney U-test, two-tailed. **f**, TH17 cells were isolated from the intestine of anti-CD3 monoclonal antibody and then restimulated *in vitro* in the presence of either anti-TGF-β+IL-6 or IL-6+TGF-β1+FICZ for 5 days. Frequencies of TR1^{exTH17} cells among total cells (left) and among exTH17 cells (right) are reported. Results are cumulative from three independent experiments. Each dot represents a pool of TH17 cells isolated from five mice treated with anti-CD3. Mean and s.e.m., *$P \le 0.05$, by Mann–Whitney U-test, two tailed.

LETTER

Hypoxia fate mapping identifies cycling cardiomyocytes in the adult heart

Wataru Kimura[1,2], Feng Xiao[1], Diana C. Canseco[1], Shalini Muralidhar[1], SuWannee Thet[1], Helen M. Zhang[3], Yezan Abderrahman[1], Rui Chen[1], Joseph A. Garcia[1,4], John M. Shelton[1], James A. Richardson[5,6], Abdelrahman M. Ashour[1], Aroumougame Asaithamby[7], Hanquan Liang[8], Chao Xing[8], Zhigang Lu[3], Cheng Cheng Zhang[3] & Hesham A. Sadek[1,9]

Although the adult mammalian heart is incapable of meaningful functional recovery following substantial cardiomyocyte loss, it is now clear that modest cardiomyocyte turnover occurs in adult mouse and human hearts[1,2], mediated primarily by proliferation of pre-existing cardiomyocytes[3–5]. However, fate mapping of these cycling cardiomyocytes has not been possible thus far owing to the lack of identifiable genetic markers[6]. In several organs, stem or progenitor cells reside in relatively hypoxic microenvironments where the stabilization of the hypoxia-inducible factor 1 alpha (Hif-1α) subunit is critical for their maintenance and function[7–10]. Here we report fate mapping of hypoxic cells and their progenies by generating a transgenic mouse expressing a chimaeric protein in which the oxygen-dependent degradation (ODD) domain of Hif-1α is fused to the tamoxifen-inducible CreERT2 recombinase. In mice bearing the *creERT2-ODD* transgene driven by either the ubiquitous CAG promoter or the cardiomyocyte-specific α myosin heavy chain promoter, we identify a rare population of hypoxic cardiomyocytes that display characteristics of proliferative neonatal cardiomyocytes, such as smaller size, mononucleation and lower oxidative DNA damage. Notably, these hypoxic cardiomyocytes contributed widely to new cardiomyocyte formation in the adult heart. These results indicate that hypoxia signalling is an important hallmark of cycling cardiomyocytes, and suggest that hypoxia fate mapping can be a powerful tool for identifying cycling cells in adult mammals.

It was recently reported that new cardiomyocytes in the adult heart are derived from pre-existing cardiomyocytes[3–5], although the identity of cycling cardiomyoctes, and the mechanism of their proliferative competency remain unknown. We recently showed that the postnatal metabolic shift from anaerobic glycolysis to oxidative phosphorylation mediates cardiomyocyte cell cycle arrest through DNA damage response[11]. Therefore, we reasoned that proliferative cardiomyocytes in the adult heart are relatively hypoxic, and thus are protected from the increased oxidative stress in the postnatal heart. In order to identify and trace the lineage of hypoxic cardiomyocytes in the adult heart, we used a Cre–*loxP*-based fate-mapping strategy to genetically and irreversibly label hypoxic cells expressing the basic helix-loop-helix transcription factor Hif-1α. Hif-1α is a master regulator of hypoxic stress response[12], and is regulated primarily by post-translational modification. Ubiquitin–proteasome-mediated degradation of Hif-1α occurs under normoxic conditions following hydroxylation of two proline residues in the ODD domain, whereas in hypoxic conditions the protein is stabilized[13–15]. Therefore, we generated a transgenic mouse line which expresses a fusion protein comprised of the ODD domain

of Hif-1α and a tamoxifen-inducible CreERT2 driven by a ubiquitous CAG promoter (Fig. 1a). We crossed these *CAG-creERT2-ODD* transgenic mice with Rosa26 floxed-stop tdTomato (R26R/tdTomato) reporter mice to irreversibly label cells that contain stabilized Hif-1α with the fluorescent protein tdTomato (Fig. 1b). Following tamoxifen induction, we found that the hypoxia detector pimonidazole co-localizes with more than half of the tdTomato[+] cardiomyocytes, which confirms the hypoxic nature of the cells (Extended Data Fig. 1), although given a significant variability of pimonidazole sensitivity depending on cell types[16], it is not clear if quantitative analysis can be applied to the pimonidazole signal (see Methods section for more details). Moreover, a significant increase in tdTomato[+] cardiomyocytes and non-cardiomyocytes was observed after brief episodes of intermittent hypoxia (6% O_2, Fig. 1b), indicating that the CreERT2–ODD system is responsive to hypoxia. Importantly, leakage of tdTomato expression without tamoxifen in *CAG-creERT2-ODD;R26R/tdTomato* transgenic mice was less than 0.00084% of total cardiomyocytes (data not shown), suggesting that tamoxifen independent non-specific labelling is unlikely to hamper data interpretation.

Next we examined several phenotypic characteristics of the tdTomato[+] cardiomyocytes using the *CAG-creERT2-ODD;R26R/tdTomato* transgenic mice 1 week after tamoxifen administration (Fig. 2a). First, we found that tdTomato[+] cardiomyocytes were surrounded by a significantly smaller number of capillaries compared with tdTomato[−] cardiomyocytes, which lends evidence to the notion of the hypoxic nature of tdTomato[+] cardiomyocytes (Extended Data Fig. 3a). Moreover, we found that tdTomato[+] cardiomyocytes harbour less nuclear 8-oxoguanine compared to tdTomato[−] cardiomyocytes (n = 3 hearts, Extended Data Fig. 3b), and that transient exposure to hypoxia reduced oxidative DNA damage in cardiomyocytes (n = 3 hearts, Extended Data Fig. 3c). In addition, tdTomato[+] cardiomyocytes were significantly smaller compared to tdTomato[−] cardiomyocytes (n = 3 hearts, Fig. 2b), and were more likely to be mononucleated (n = 3 hearts, Fig. 2c).

To assess whether these tdTomato[+] hypoxic cardiomyocytes contribute to new cardiomyocyte formation in the adult heart, we administered tamoxifen to *CAG-creERT2-ODD;R26R/tdTomato* mice and traced their lineage over 4 weeks (Fig. 2a). We found that only a very small number of cardiomyocytes were labelled 3 days or 1 week after tamoxifen pulse (n = 3 hearts each, Fig. 2d). Consistent with previous reports[17–19], we did not detect c-Kit or Sca-1 expression in cardiomyocytes (including tdTomato[+] myocytes (data not shown)). Since CAG is a ubiquitous promoter, tdTomato[+] cells were also found in non-cardiomyocyte lineages, including vascular endothelial and smooth muscle cells, as well as interstitial fibroblasts

[1]Department of Internal Medicine, Division of Cardiology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [2]Life Science Center, Tsukuba Advanced Research Alliance, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8577, Japan. [3]Departments of Physiology and Developmental Biology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [4]Department of Medicine, VA North Texas Health Care System, 4600 South Lancaster Road, Dallas, Texas 75216, USA. [5]Department of Molecular Biology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [6]Department of Pathology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [7]Department of Radiation Oncology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [8]McDermott Center for Human Growth and Development, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [9]Hamon Center for Regenerative Science and Medicine, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.
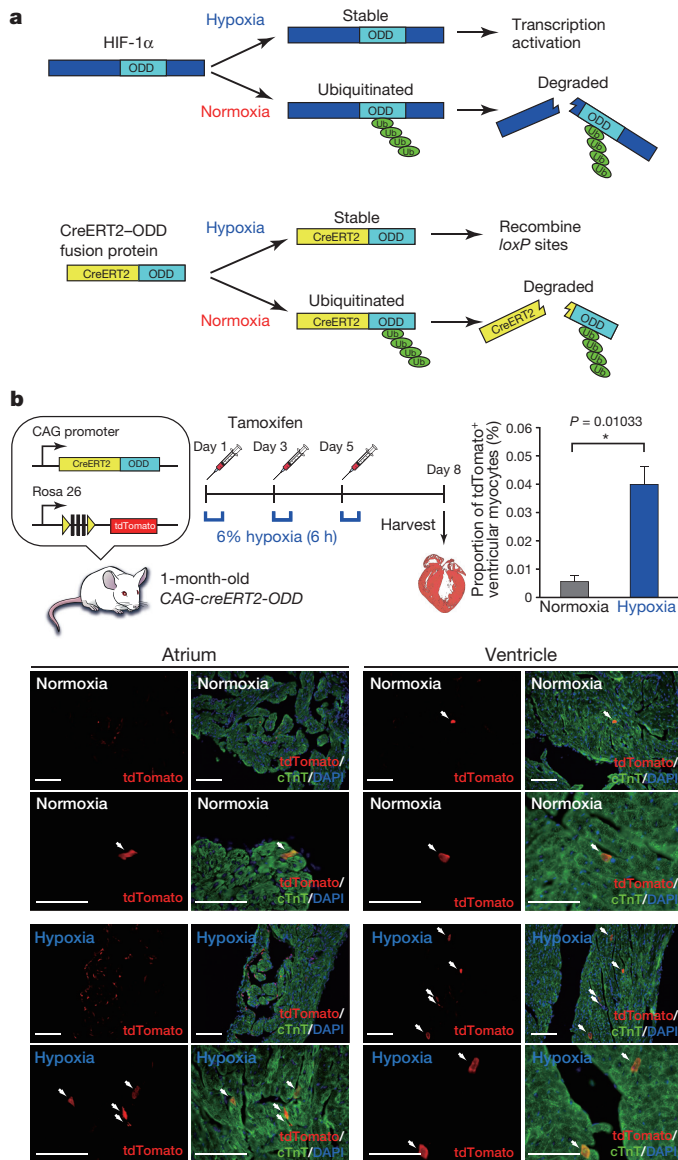
Figure 1 | Generation of Hif-1α-dependent fate-mapping model.
a, Schematic diagram of a strategy to activate Cre protein in a Hif-1α-stability-dependent manner. Under normoxic conditions, the ODD domain of Hif-1α is hydroxylated and ubiquitylated leading to E3 ubiquitin–proteasome-dependent protein degradation, whereas under hypoxic conditions, the ODD domain is not hydroxylated and avoids protein degradation. In order to activate Cre in hypoxic conditions, we generated transgenic mice in which the CAG-promoter-driven CreERT2–ODD fusion protein is expressed so that the CreERT2 protein is only stabilized in Hif-1α-expressing hypoxic cells.
b, Validation of the hypoxia-dependent model was performed by exposing mice to intermittent hypoxia (6% O_2 for 6 h every other day) at the same time as tamoxifen pulse, before hearts were harvested 3 days later. Images and graph show a significant increase in the number of both atrial and ventricular cardiomyocytes with tdTomato fluorescence after hypoxia exposure in *CAG-creERT2-ODD;R26R/tdTomato* transgenic hearts. Scale bars indicate 100 μm. *$P < 0.05$, **$P < 0.01$. A two-tailed unpaired *t*-test was used for statistical analysis.

1 week after tamoxifen administration (Extended Data Fig. 4) Importantly, we did not observe overlap of c-Kit[+] or Sca-1[+] cells with tdTomato[+] cells (data not shown). The endothelial, smooth muscle and fibroblast lineages showed temporal expansion at 4 weeks following tamoxifen administration (Extended Data Fig. 4). Notably, 4 weeks after tamoxifen administration, we also observed a significant increase in the number of tdTomato[+] cardiomyocytes both in atria and ventricles compared to earlier time points, which indicates

that these hypoxic cardiomyocytes contribute to cardiomyocyte turnover in the adult heart ($n = 3$ hearts, Fig. 2d). The rate of new cardiomyocyte formation by tdTomato[+] ventricular cardiomyocytes was $0.9781 \pm 0.09790\%$ per year, which is similar to the rate of myocyte turnover seen in previous reports[3]. Of note, $39.94 \pm 5.268\%$ of tdTomato[+] cardiomyocytes formed clusters at 1 month after tamoxifen pulse ($n = 3$ hearts, Fig. 2e), suggesting clonal expansion of hypoxic cardiomyocytes. To assess whether the increase in number of tdTomato[+] cardiomyocytes is the result of cell fusion or bona fide cardiomyocyte cell division, we used the Rosa 26 floxed-stop mTmG (R26R/mTmG) double colour reporter line (see Methods for details). We found that the increase in the number of eGFP[+]/tdTomato[−] cardiomyocytes between 1 and 4 weeks was statistically significant, whereas the rate of eGFP[+]/tdTomato[+] (fusion) cardiomyocytes did not change ($n = 3$ hearts, Extended Data Fig. 3d). This indicates that the newly labelled cardiomyocytes resulted from proliferation rather than cell fusion events.

Given that the CAG promoter drives ubiquitous expression in all lineages, it is difficult to exclude the possibility that tdTomato[+] non-cardiomyocytes, such as stem or progenitor cells, contributed to new cardiomyocyte formation. Therefore, we generated another transgenic mouse line harbouring the *creERT2-ODD* fusion transgene driven by cardiomyocyte-specific α myosin heavy chain (αMHC) promoter. As a proof of concept, we first show that cardiomyocytes that are positive for immunostaining using an anti-Cre antibody were also positive for Hif-1α (Extended Data Fig. 5). Next we crossed the αMHC-CreERT2–ODD and the R26R/tdTomato reporter lines. This allowed us to identify a rare population of cardiomyocytes ($0.051 \pm 0.0075\%$ of total cardiomyocytes) that were labelled with tdTomato at 1 week after tamoxifen administration ($n = 5$ hearts), which is consistent with our earlier results using the CAG line, except that no non-cardiomyocytes were labelled at any time point in any of the mice ($n = 27$, data not shown). Leakage of tdTomato[+] cardiomyocytes without tamoxifen administration was less than 0.005% (data not shown). Interestingly, tdTomato[+] cardiomyocytes showed significantly lower signal intensity of immunofluorescence with an antibody recognizing hydroxylated Hif-1α proline 402 residue ($n = 5$ hearts, Fig. 3a), which is an indicator of hypoxic stabilization of Hif-1α[20]. In addition, quantification of the number of capillaries surrounding tdTomato[+] cardiomyocytes revealed significantly lower capillary density surrounding hypoxic cardiomyocytes, which is also consistent with our earlier results using the CAG line, and is indicative of their hypoxic environment ($n = 5$ hearts, Fig. 3b). Furthermore, we found that hypoxic cardiomyocytes had significantly lower levels of oxidative DNA damage ($n = 5$ hearts, Fig. 3c), smaller cell size ($n = 5$ hearts, Fig. 3d), and were more likely to be mononucleated ($n = 5$ hearts, Fig. 3e), all which are features of proliferative fetal/neonatal cardiomyocytes.

To further examine the molecular phenotype of hypoxic cardiomyocytes in their native environment, we isolated tdTomato[+] cardiomyocytes from fresh cryosections using laser microdissection ($n = 2$ hearts, 44 cells, Extended Data Fig. 6a). A comprehensive analysis of gene expression using RNA sequencing (RNA-seq) revealed downregulation of negative regulators of Hif-1α (Extended Data Fig. 6b) as well as upregulation of positive regulators of Hif-1α (Extended Data Fig. 6c) and Hif-1α target genes (Extended Data Fig. 6d) in hypoxic cardiomyocytes. RNA-seq analysis also showed the upregulation of CDK/cyclins (Extended Data Fig. 6e) along with the downregulation of negative cell cycle regulators such as CDK inhibitors, cell cycle checkpoint genes, and DNA damage response genes (Extended Data Fig. 6f) in hypoxic cardiomyocytes. Notably, we found that several Hif genes including Hif-1α and Hif-1β are upregulated in hypoxic cardiomyocytes (Extended Data Fig. 6g). Importantly, we also found that the negative cardiomyocyte cell cycle regulator Meis1 and other Meis family members (Extended Data Fig. 6h), as well as markers of cardiomyocyte hypertrophy (Extended Data Fig. 6i), were all
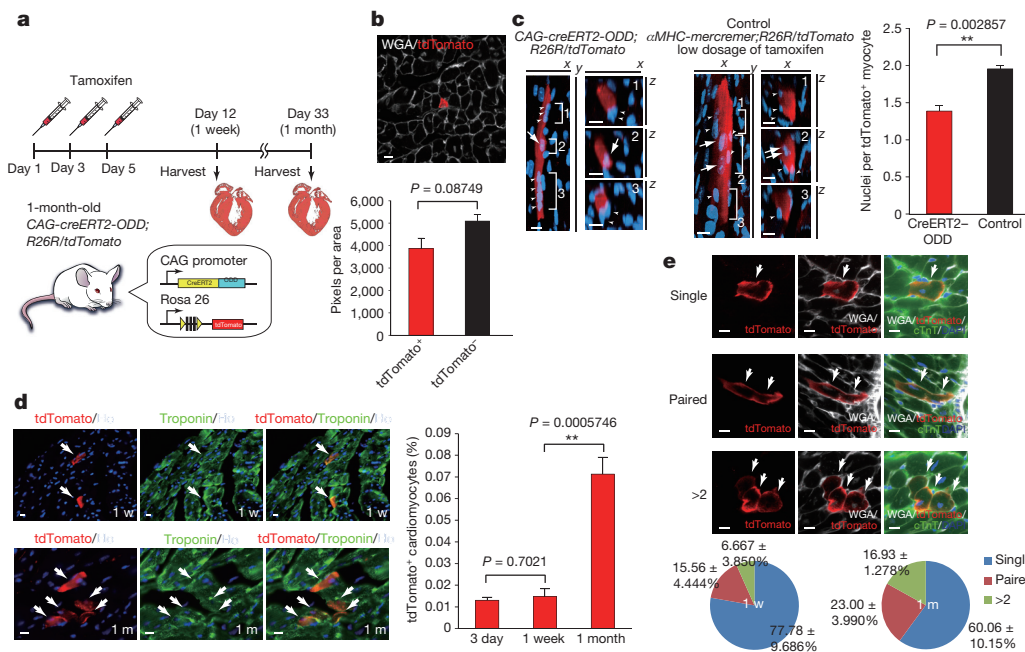
Figure 2 | Hypoxic cells contribute to cardiomyocyte turnover in the young adult heart. a, Schematic diagram of the lineage tracing experiment. **b**, Cell size of tdTomato$^+$ hypoxic cardiomyocytes at 1 week after tamoxifen pulse is significantly smaller than that of surrounding non-labelled cardiomyocytes. WGA, wheat germ agglutinin. **c**, Almost half of tdTomato$^+$ hypoxic cardiomyocytes were mononucleated, whereas control cardiomyocytes were, on average, binucleated. Representative confocal *z*-stack images of tdTomato$^+$ cardiomyocytes in CAG-CreERT2–ODD;R26R/tdTomato and control αMHC-merCremer;R26R/tdTomato mice are shown. Arrows indicate nuclei belonging to tdTomato-labelled cardiomyocytes, judged by the fact that the DAPI signal is completely surrounded by tdTomato signal.

Arrowheads indicate nuclei outside of the labelled cardiomyocyte. **d**, Three days or one week after tamoxifen pulse at 1 month of age, a few cardiomyocytes were labelled with tdTomato (indicated by arrows). A substantial increase in number of tdTomato$^+$ cardiomyocytes is observed over 1 month. Nuclei are stained with Hoechst 33258 (Ho). **e**, Wheat germ agglutinin co-staining showed clusters of tdTomato-labelled cardiomyocytes at 1 week after tamoxifen pulse followed by an increase in the rate of clustered cardiomyocytes over 1 month, indicating clonal expansion of cardiomyocytes. Data are presented as mean ± s.e.m. *$P < 0.05$, **$P < 0.01$. A two-tailed unpaired *t*-test was used for statistical analysis. Scale bars represent 10 μm.

significantly downregulated in hypoxic cardiomyocytes. Ingenuity pathway analysis showed enrichment of differentially expressed genes from several pathways related to cell cycle progression, oxidative stress response, and DNA repair (heat maps of these pathways are shown in Extended Data Fig. 7) in hypoxic cardiomyocytes. These results not only support the hypoxic/proliferative nature of cycling cardiomyocytes, but they also indicate that there are additional intrinsic regulatory mechanisms that favour Hif-1α expression and stabilization.

Next, we traced the lineages of hypoxic cardiomyocytes in the *αMHC-creERT2-ODD;R26R/tdTomato* transgenic line at 1 or 2 months of age, and for 1 month (*n* = 5 hearts for 1-month-old, *n* = 4 hearts for 2-month-old) and 2 months (*n* = 4 hearts) (Fig. 4a) following tamoxifen administration. We observed that the small population of cardiomyocytes initially labelled after tamoxifen administration significantly increased within 1 month, followed by further expansion at 2 months (Fig. 4b), with an annual rate of cardiomyocyte formation of 0.6219 ± 0.1319%. Both the initial localization of hypoxic cardiomyocytes as well as their progeny were uniformly distributed throughout the myocardium (Extended Data Fig. 9). Visualization of cell–cell boundaries with wheat germ agglutinin antibody staining demonstrated that 37.43 ± 11.09% of tdTomato$^+$ cardiomyocytes form clusters comprised of two or more cells (Extended Data Fig. 8a), providing additional support for the clonal expansion of tdTomato$^+$ cardiomyocytes. Moreover, tdTomato$^+$ cardiomyocytes showed a significant increase in Ki67 and 5-bromo-2′-deoxyuridine (BrdU) labelling compared with tdTomato$^-$ cardiomyocytes. In fact, the vast majority of BrdU$^+$ cardiomyoctes were derived from tdTomato$^+$ cardiomyocytes (Fig. 4c and Extended Data Fig. 8b). In addition, myocardial infarction injury induced by permanent ligation of the

left anterior descending coronary artery, which is known to accelerate cardiomyocyte turnover[3], resulted in a significant increase in the number of tdTomato$^+$ cardiomyocytes and in BrdU incorporation (*n* = 3 hearts, Extended Data Fig. 8c). Moreover, lineage tracing of αMHC-CreERT2–ODD;R26R/mTmG mice for 1 week (*n* = 5 hearts) or 1 month (*n* = 6 hearts) after tamoxifen administration (Extended Data Fig. 10a) indicated that expansion of labelled cardiomyocytes is mainly through bona fide cell proliferation rather than cellular fusion (Extended Data Fig. 10b) with no significant difference in the rate of fusion events in different myocardial regions (Extended Data Fig. 10c). In summary, these results indicate that a rare population of hypoxic cardiomyocytes with immature phenotypic characteristics contributes widely to new cardiomyocyte formation in the adult heart.

Our Cre–*loxP* genetic lineage tracing system, using post-translational protein stability instead of gene expression, enabled us to identify and fate map a previously unidentified rare population of cycling hypoxic cardiomyocytes in the adult heart. Although we provide several lines of evidence that support the hypoxic nature of these cycling cardiomyocytes, such as lower capillary density and co-localization with the hypoxia marker pimonidazole, our RNA-seq analysis suggests that other endogenous mechanisms, Hif-1α stabilization, and maintenance of hypoxia signalling are involved. In particular, the upregulation of Hif-1 mRNA, and the downregulation of prolyl hydroxylases indicates that these cycling cardiomyocytes are intrinsically programmed to maintain hypoxia signalling, although the mechanism of this endogenous regulation is unclear.

These data suggest that hypoxic cardiomyocytes contribute to new cardiomyocyte formation in the adult heart at a rate of ~0.3–1% annually, which is well within previously reported rates of cardiomyocyte turnover[1,3,21]. Even though the majority of BrdU- or Ki67-positive
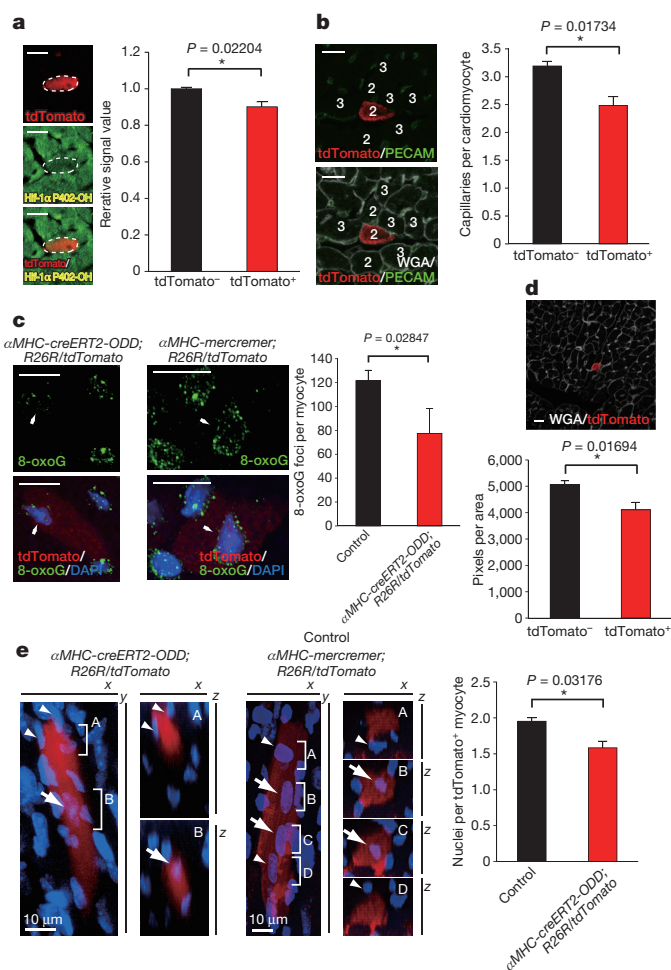
Figure 3 | Hypoxic cardiomyocytes share characteristics of proliferative cardiomyocytes. **a**, Immunofluorescence and measurement of signal intensity demonstrated less hydroxylation of the proline 402 residue of Hif-1α (Hif-1α P402-OH), indicative of relative hypoxia and the stabilization of the Hif-1α protein. **b**, tdTomato-labelled cardiomyocytes in αMHC-CreERT2-ODD;R26R/tdTomato mice at 1 week after tamoxifen pulse were juxtaposed to fewer capillary blood vessels compared with surrounding non-labelled cardiomyocytes. **c**, Oxidative DNA damage, as indicated by the quantification of nuclear foci of an anti-8-oxoguanine (8-oxoG) antibody, was significantly lower in tdTomato+ cardiomyocytes compared with surrounding non-labelled cardiomyocytes at 1 week after tamoxifen pulse. **d**, Cell size of tdTomato+ hypoxic cardiomyocytes at 1 week after tamoxifen pulse is significantly smaller than that of surrounding non-labelled cardiomyocytes. **e**, Almost half of tdTomato+ cardiomyocytes were mononucleated, whereas control cardiomyocytes were, on average, binucleated. Representative confocal z-stack images of tdTomato+ cardiomyocytes in αMHC-CreERT2-ODD;R26R/tdTomato and control αMHC-mercremer;R26R/tdTomato mice are shown. Arrows indicate nuclei belong to tdTomato-labelled cardiomyocytes, judged by the fact that the DAPI signal is completely surrounded by tdTomato signal. Arrowheads indicate nuclei outside of the labelled cardiomyocyte. A–D indicate the level of optical sections shown on right. *P < 0.05, **P < 0.01. A two-tailed unpaired t-test was used for statistical analysis. Scale bars indicate 10 μm.

cardiomyocytes originate from the hypoxic cardiomyocyte population, both at baseline and following ischaemic injury, it is possible that there are minor contributions to new cardiomyocyte formation from other progenitor cells or cardiomyocyte populations not labelled by our fate-mapping approach.

Recently, it has become increasingly evident that hypoxia signalling is critical for maintenance of proliferative competency of numerous stem and progenitor populations, possibly as a default pathway for avoiding
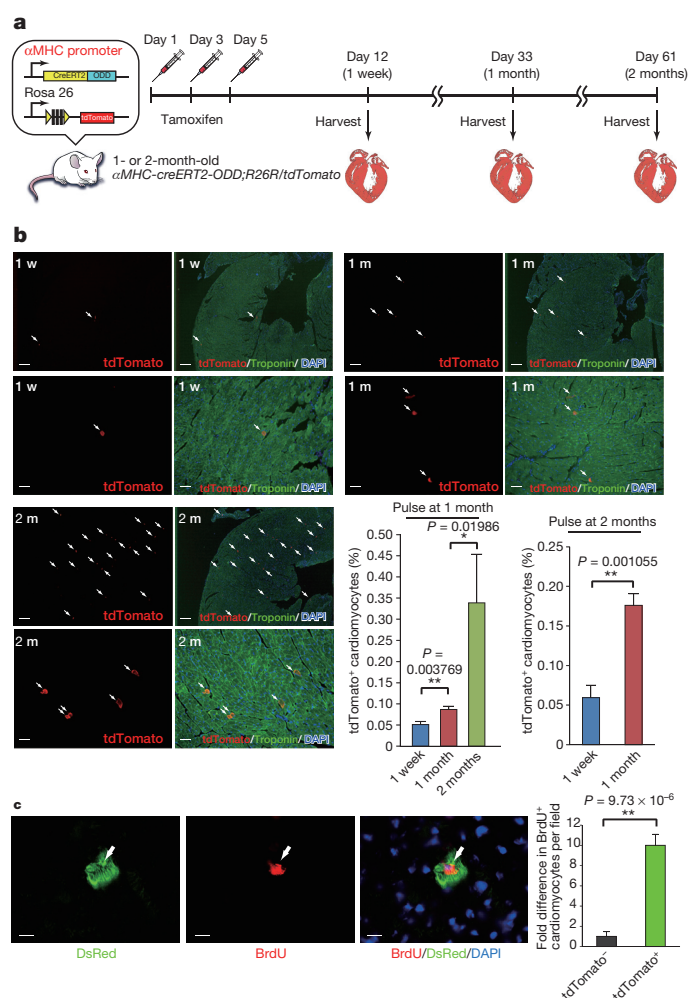
Figure 4 | Hypoxic cardiomyocytes are cycling in the adult hearts.
a, Schematic diagram of the lineage tracing experiment. **b**, One week after tamoxifen pulse (1 w) at both 1 and 2 months (m) of age, few cardiomyocytes were labelled with tdTomato (indicated by arrows). A substantial increase in the number of tdTomato+ cardiomyocytes was observed 1 month (1M) after tamoxifen pulse. The number increased further 2 months after tamoxifen pulse (2M). Scale bars represent 200 μm. **c**, Co-immunostaining with anti-BrdU and anti-DsRed antibodies showed a significantly increased rate of BrdU incorporation in tdTomato+ cardiomyocytes after 1 month of BrdU administration started at the time of tamoxifen administration. *P < 0.05, **P < 0.01. A two-tailed unpaired t-test was used for statistical analysis. Scale bars represent 10 μm.

oxidative stress and maintenance of genomic integrity[22–24]. As such, the hypoxia fate-mapping strategy outlined in the current study can provide important clues about the role of hypoxia signalling in cellular turnover.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Bergmann, O. et al. Evidence for cardiomyocyte renewal in humans. Science **324**, 98–102 (2009).
2. Malliaras, K. et al. Cardiomyocyte proliferation and progenitor cell recruitment underlie therapeutic regeneration after myocardial infarction in the adult mouse heart. EMBO Mol. Med. **5**, 191–209 (2013).
3. Senyo, S. E. et al. Mammalian heart renewal by pre-existing cardiomyocytes. Nature **493**, 433–436 (2013).
4. Bersell, K., Arab, S., Haring, B. & Kuhn, B. Neuregulin1/ErbB4 signaling induces cardiomyocyte proliferation and repair of heart injury. Cell **138**, 257–270 (2009).

5. Ali, S. R. *et al.* Existing cardiomyocytes generate cardiomyocytes at a low rate after birth in mice. *Proc. Natl Acad. Sci. USA* **111,** 8850–8855 (2014).
6. Parmacek, M. S. & Epstein, J. A. Cardiomyocyte renewal. *N. Engl. J. Med.* **361,** 86–88 (2009).
7. Mazumdar, J. *et al.* O_2 regulates stem cells through Wnt/β-catenin signalling. *Nature Cell Biol.* **12,** 1007–1013 (2010).
8. Takubo, K. *et al.* Regulation of the HIF-1α level is essential for hematopoietic stem cells. *Cell Stem Cell* **7,** 391–402 (2010).
9. Culver, J. C., Vadakkan, T. J. & Dickinson, M. E. A specialized microvascular domain in the mouse neural stem cell niche. *PLoS ONE* **8,** e53546 (2013).
10. Parmar, K., Mauch, P., Vergilio, J. A., Sackstein, R. & Down, J. D. Distribution of hematopoietic stem cells in the bone marrow according to regional hypoxia. *Proc. Natl Acad. Sci. USA* **104,** 5431–5436 (2007).
11. Puente, B. N. *et al.* The oxygen-rich postnatal environment induces cardiomyocyte cell-cycle arrest through DNA damage response. *Cell* **157,** 565–579 (2014).
12. Semenza, G. L. Oxygen homeostasis. *Wiley Interdisc. Rev. Syst. Biol. Med.* **2,** 336–361 (2010).
13. Tang, Y. L. *et al.* A hypoxia-inducible vigilant vector system for activating therapeutic genes in ischemia. *Gene Ther.* **12,** 1163–1170 (2005).
14. Ivan, M. *et al.* HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O_2 sensing. *Science* **292,** 464–468 (2001).
15. Jaakkola, P. *et al.* Targeting of HIF-α to the von Hippel–Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation. *Science* **292,** 468–472 (2001).
16. Sato, Y. *et al.* Cellular hypoxia of pancreatic β-cells due to high levels of oxygen consumption for insulin secretion *in vitro. J. Biol. Chem.* **286,** 12524–12532 (2011).
17. Sanada, F. *et al.* c-Kit-positive cardiac stem cells nested in hypoxic niches are activated by stem cell factor reversing the aging myopathy. *Circ. Res.* **114,** 41–55 (2014).
18. van Berlo, J. H. *et al.* c-kit$^+$ cells minimally contribute cardiomyocytes to the heart. *Nature* **509,** 337–341 (2014).
19. Uchida, S. *et al.* Sca1-derived cells are a source of myocardial renewal in the murine adult heart. *Stem Cell Rep.* **1,** 397–410 (2013).
20. Tian, Y. M. *et al.* Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. *J. Biol. Chem.* **286,** 13041–13051 (2011).
21. Soonpaa, M. H., Kim, K. K., Pajak, L., Franklin, M. & Field, L. J. Cardiomyocyte DNA synthesis and binucleation during murine development. *Am. J. Physiol.* **271,** H2183–H2189 (1996).
22. Ito, K. & Suda, T. Metabolic requirements for the maintenance of self-renewing stem cells. *Nature Rev. Mol. Cell Biol.* **15,** 243–256 (2014).
23. Zhang, C. C. & Sadek, H. A. Hypoxia and metabolic properties of hematopoietic stem cells. *Antioxid. Redox Signal.* **20,** 1891–1901 (2014).
24. Jopling, C., Sune, G., Faucherre, A., Fabregat, C. & Izpisua Belmonte, J. C. Hypoxia induces myocardial regeneration in zebrafish. *Circulation* **126,** 3017–3027 (2012).

METHODS

Generation of transgenic mice. The CAG-CreERT2–ODD vector was assembled with three-piece ligation including (1) 3.2 kb fragment of pCAG-CreERT2 (Addgene plasmid #14797) cut with XhoI and NotI, (2) PCR product with forward (5′-GATGGCGATCTCGAGCCATC-3′) and reverse (5′-ACGTGGTACCAGCTGTGGCAGGGAAACCCT-3′) primers using pCAG-CreERT2 as a template and cut with XhoI and KpnI, and (3) the PCR product with forward (5′-ACGTGGTACCAAGTTGGAATTGGTAGAAAAACTTTTTGC-3′) and reverse (5′-TCACGGGCCGCTCAGGCGTCTTCAGTAGTTTCTTTATG-3′) primers using pR26-ODD-luc (provided by W. Kaelin[25]) as a template and cut with KpnI/NotI. The vector was linearized with SpeI and injected into fertilized eggs to generate CAG-CreERT2–ODD transgenic mice. Founder (F0) mice were identified by PCR using genomic DNA isolated from the tails with forward (5′-TCCTCTCCCACATCAGGCAC-3′) and reverse (5′-TGAACCAGCTCCCTGTCTGC-3′) primers. These primer set were also used for genotyping for the following experiments. To generate a targeting vector containing the cardiomyocyte-specific αMHC promoter followed by the cDNA of CreERT2–ODD fusion protein, we isolated BamHI-digested αMHC promoter taken from αMHC-eGFP-Rex-Neo plasmid (Addgene Plasmid #21229) and PCR-amplified CreERT2–ODD fragment using the following primers: forward, 5′-CAGGCGACTAGTATGTCCAATTTACTGAAC-3′; reverse, 5′-CTTTTTGAGCTCCTGCAGGTCGAGGGATCT-3′ using a CAG-CreERT2–ODD-targeting vector as a template. We then inserted these two fragments into pBluescript SK+ vector followed by linearization with SalI digestion and gel purification for the injection into fertilized eggs of B6/C3H chimaera. The primer set for the CAG-CreERT2–ODD line was used for the genotyping and maintenance of the αMHC-creERT2-ODD transgenic line. Transgenic F0 mice were crossed with C57BL6/J mice, and transgenic F1 mice were crossed with R26R-tdTomato and R26R-mTmG.

Animals. All protocols were approved by the Institutional Animal Care and Use Committee of the University of Texas Southwestern Medical Center. All experiments were performed on age-matched mice with equal ratio of male to female. Rosa26 reporter tdTomato (R26R-tdTomato, B6;129S6-Gt(ROSA)26Sor^tm9(CAG-tdTomato) Hze/J)), Rosa26 reporter mTmG (R26R/mTmG, B6.129(Cg)-Gt(ROSA)26 Sor^tm4(ACTB-tdTomato,-EGFP)Luo/J) obtained from the Jackson Laboratory. All experiments were performed on age- and sex-matched mice with equal ratio of male and female mice. Healthy mice were chosen randomly from the expansion colony for experiment. Transgenic mice were maintained on a mixed genetic background of 50% B6 and 50% 129 because we observed spontaneous silencing of transgene in transgenic mice backcrossed to B6 and also C3H. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. CAG- or αMHC-creERT2-ODD;R26/ mTmG double transgenic mice were used to assess cell fusion. In this line, membrane-targeted tdTomato is expressed under the control of ubiquitous promoter on Rosa26 locus, whereas membrane-targeted eGFP becomes active after Cre-mediated excision of floxed tdTomato cDNA, and therefore eGFP$^+$/tdTomato$^+$ cells are indicative of cell fusion between eGFP$^+$/tdTomato$^-$ hypoxic cells and eGFP-/tdTomato$^+$ normoxic cells.

Genotyping. R26R-tdTomato and R26R-mTmG mice were genotyped by PCR with tail DNA as described in the Jackson Laboratory Genotyping Protocols. Primer sequences are as follows: R26R-tdTomato, wild type forward, 5′-AA GGGAGCTGCAGTGGAGTA-3′, wild type reverse, 5′-CCGAAAATCTGTG GGAAGTC-3′, mutant forward, 5′-CTGTTCCTGTACGGCATGG-3′, and mutant reverse, 5′-GGCATTAAAGCAGCGTATCC-3′; R26R-mTmG, wild type forward, 5′-CTCTGCTGCCTCCTGGCTTCT-3′, wild type reverse, 5′-CGAGGCGGATCA CAAGCAATA-3′, and mutant reverse, 5′-TCAATGGGCGGGGGTCGTT-3′.

Drug administration. Tamoxifen (Sigma) was dissolved in 90% sesame oil (Sigma)/10% ethanol and stored at −20°C. Prior to intraperitoneal injection (1 mg per day per mouse), the tamoxifen solution was heated at 55°C for 10 min. 5-bromo-2′-deoxyuridine (BrdU, MP biomedical) was introduced in the drinking water for 4 weeks after tamoxifen pulse, or for 1 week after coronary artery ligation before harvesting for cryosections. Pimonidazole HCl (60 mg kg^{-1}, dissolved in PBS) was injected into the tail vein and tissues were harvested 5 min, 15 min, 90 min and 180 min after injection. It is important to note here that we found some pimonidazole-positive nucleated blood cells in the ventricular chambers (data not shown). Given the high oxygen tension in the circulation, this suggests that there may be additional cell-specific factors that mediate pimonidazole detection in hypoxic cells, as has been previously reported in the literature[16].

Laser microdissection and RNA-seq. Tamoxifen was injected to αMHC-CreERT2–ODD;R26R/tdTomato mice once a day for 3 days, then the hearts were harvested and embedded in freezing medium without fixation. Eight-micrometre sections were mounted on PPS membrane frame slides (Leica), and tdTomato$^+$ cardiomyocytes were collected with AS-LMD laser microdissection system (Leica) controlled with LMD v5.0C software (Leica). RNA was purified from dissected

cardiomyocytes with Qiagen RNeasy Mini Kit according to the manufacturer's instruction, reverse-transcribed, and amplified following published procedures[26]. The cDNA was sonicated using the Covaris S2 ultrasonicator, and libraries were prepared with the KAPA High Throughput Library Preparation Kit. Samples were end-repaired, 3′ ends adenylated and barcoded with multiplex adapters. PCR-amplified libraries were purified with AmpureXP beads, and validated on the Agilent 2100 Bioanalyzer. Before being normalized and pooled, samples were quantified by Qubit (Invitrogen) and then run on a Illumina Hiseq 2500 using PE100 SBS v3 reagents to generate 51-bp single-end reads. Before mapping, reads were trimmed to remove low-quality regions in the ends. Trimmed reads were mapped to the mouse genome (mm10) using TopHat v2.0.12[27] with the UCSC iGenomes GTF file from Illumina. Alignments with mapping quality less than 10 were discarded. Expression abundance estimation and differential expression gene identification was done using edgeR[28]. Genes with $\log_2$(fold change)> 2 and FDR <0.05 were deemed significantly differentially expressed between the two conditions. Pathway analysis was conducted using QIAGEN's Ingenuity Pathway Analysis tool (QIAGEN Redwood City, http://www.qiagen.com/ingenuity). Differentially expressed gene heat maps were clustered by hierarchical clustering (hclust function in R, http://www.R-project.org).

Immunofluorescence. Tissues were fixed in 4% paraformaldehyde (PFA)/PBS for 1 h at room temperature and then incubated in 30% sucrose/PBS at 4°C until tissues sunk. Tissues were embedded in freezing medium, frozen at −80°C and cut immediately. Sections were incubated at 4°C overnight with primary antibodies after washing with PBS followed by blocking with 3% serum from the host animal of secondary antibody/0.3% Triton-X100/PBS for 1 h at room temperature. A monoclonal antibody against pimonidazole (FITC-MAb1, clone 4.3.11.3, Hypoxiprobe-1 Plus Kit) was used to visualize pimonidazole. Despite an intensive effort to optimize the staining condition, the signal of pimonidazole antibody was not uniform and therefore concluded not to be usable for quantification. For the staining with anti-8-oxoguanine (8-oxoG) and anti-DsRed antibodies, sections were fixed in 4% PFA for 20 min at room temperature, washed with PBS and boiled for 40 min in 1 mM EDTA/0.05% Tween-20 for antigen retrieval, and then blocked and incubated with primary antibodies overnight at room temperature. After washing with PBS, sections were incubated with secondary antibodies (Invitrogen) for 1 h at room temperature, washed with PBS, stained with DAPI and mounted with VECTASHIELD for imaging. For staining 100-mm-thick sections, floating free heart sections were washed with 0.1% Triton-X100/PBS and then blocked with 3% serum/ 0.1% Triton-X100/PBS for 3 h at room temperature, and then incubated with primary antibodies at 4°C for 24 h in blocking solution. Sections were then washed and incubated with secondary antibodies for 3 h at room temperature. Sections were then cover-slipped after DAPI staining.

Antibodies. Primary antibodies and dilutions are following: anti-troponin T, cardiac isoform Ab-1, clone 13-11 (Thermo scientific MS-295-P1, 1:100), anti-Cre (Novus Biologicals NB100-56133, 1:100 or Santa Cruz Biotechnology sc-83398, 1:100), anti-sarcomeric α-actinin (Abcam ab68167, 1:100), anti-green fluorescent protein (GFP) antibody (Aves GFP-1010, 1:400), anti-8-oxoG (Abcam ab64548, 1:100), anti-DsRed (Clontech #632496, 1:400), anti-wheat germ agglutinin (WGA) conjugated with AlexaFluor 647 (Invitrogen, 20 μg ml^{-1}), anti-bromodeoxyuridine (Roche 11170376001, 1:25), anti-PECAM (PharMingen #553370, 1:20), anti-SM22α (Abcam ab14106, 1:100), anti-vimentin (Abcam ab28028, 1:100), anti-CA9 (R&D systems AF2344, 1:100), anti-Hif-1α (Santa Cruz Biotechnology sc-10790, 1:100), anti-HIF-1α (hydroxy P402) (Abcam ab72775), anti-Ki67 (eBioscience, 14-5698-82, 1:50).

Imaging. Fluorescent tissue images were obtained with Leica DM2000 or Zeiss Axio Scan microscopes. Images taken with Axio Scan were processed with Photoshop CS2 and ZEN 2012 to generate merge images with different colours. Confocal images were obtained with Zeiss LSM510 microscope and processed with AutoQuant for 3D deconvolution and with Imaris to reconstruct 3D images from multiple z-stack recordings. To count total number of cardiomyocytes in ventricles, images of WGA-stained sections were taken with AxioScan, processed with Photoshop, and then analysed with ImageJ. The image was then manually examined to verify the automated counting (Extended Data Fig. 2). For the analysis of cardiomyocyte nucleation, 100-μm-thick cryosections were analysed with LSM510 confocal microscope, and only cardiomyocytes with a complete cell body, as confirmed by WGA staining, underwent nucleation analysis. For the quantification of nuclear 8-oxoG foci, cryosections were scanned with LSM510 confocal microscope and analysed with Imaris software after deconvolution with AutoQuant software[11]. For both nucleation and 8-oxoG foci quantification, we used a αMHC-merCremer;R26R/tdTomato mouse line as control to ensure uniformity of Cre recombination and reporter expression between the two groups.

Data analysis. For quantification of the number of fluorescent protein labelled cardiomyocyte, the results acquired from at least three to five sections of the heart

harvested from each animal at the ventricular valve level of the four-chamber view with at least 100 μm distance from each other were averaged. All graphs represent average values, and all error bars represent s.e.m. No statistical methods were used to predetermine sample size. All data collected and analysed were assumed to be distributed normally. The two-tailed unpaired student t-test was used to determine statistical significance and $P < 0.05$ was considered statistically different.

Quantification of capillary number. The number of capillaries surrounding each cardiomyocyte was counted by visualizing cell membrane and capillary endothelial cells with immunostaining with anti-WGA antibody and anti-PECAM antibody, respectively on cryosections with 8 μm thickness, which is an established method to assess capillary density of cardiomyocytes[29].

25. Safran, M. *et al.* Mouse model for noninvasive imaging of HIF prolyl hydroxylase activity: assessment of an oral agent that stimulates erythropoietin production. *Proc. Natl Acad. Sci. USA* **103,** 105–110 (2006).
26. Tang, F. *et al.* RNA-seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols* **5,** 516–535 (2010).
27. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25,** 1105–1111 (2009).
28. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
29. Johansson, B., Morner, S., Waldenstrom, A. & Stal, P. Myocardial capillary supply is limited in hypertrophic cardiomyopathy: a morphological analysis. *Int. J. Cardiol.* **126,** 252–257 (2008).
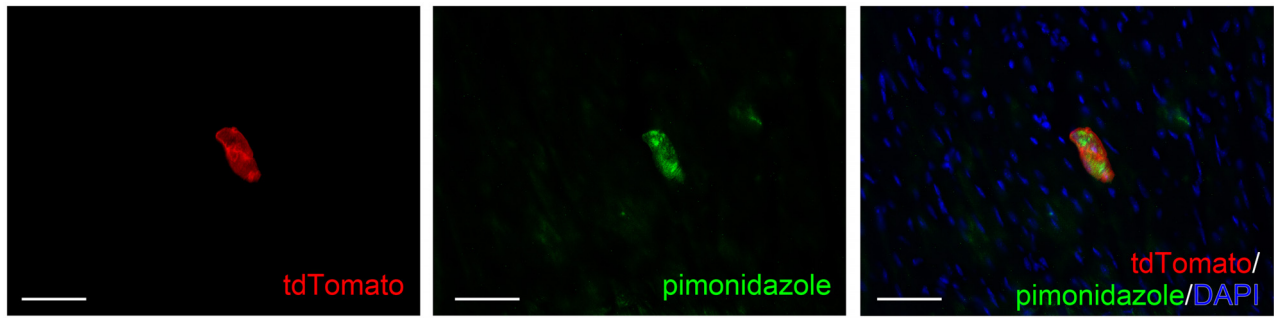
Extended Data Figure 1 | tdTomato⁺ cardiomyocytes in *creERT2-ODD;R26R/tdTomato* transgenice mice were pimonidazole⁺ at 3 days after tamoxifen administration. Scale bars indicate 50 μm.
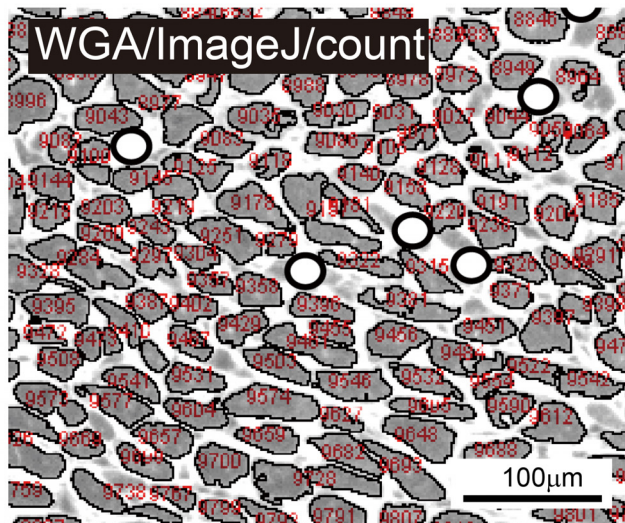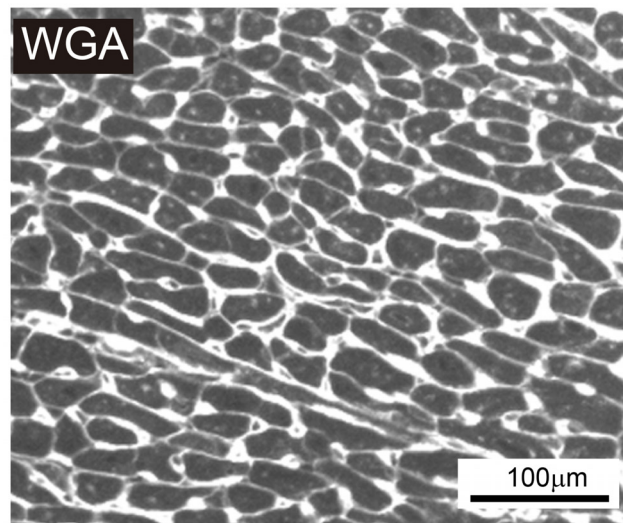
Extended Data Figure 2 | Counting total number of cardiomyocyte included in four-chamber view section at the level of aortic valve. Sections stained with anti-WGA was analysed with ImageJ the number of cardiomyocytes in ventricles were manually counted. Circles indicate manually counted cardiomyocytes.
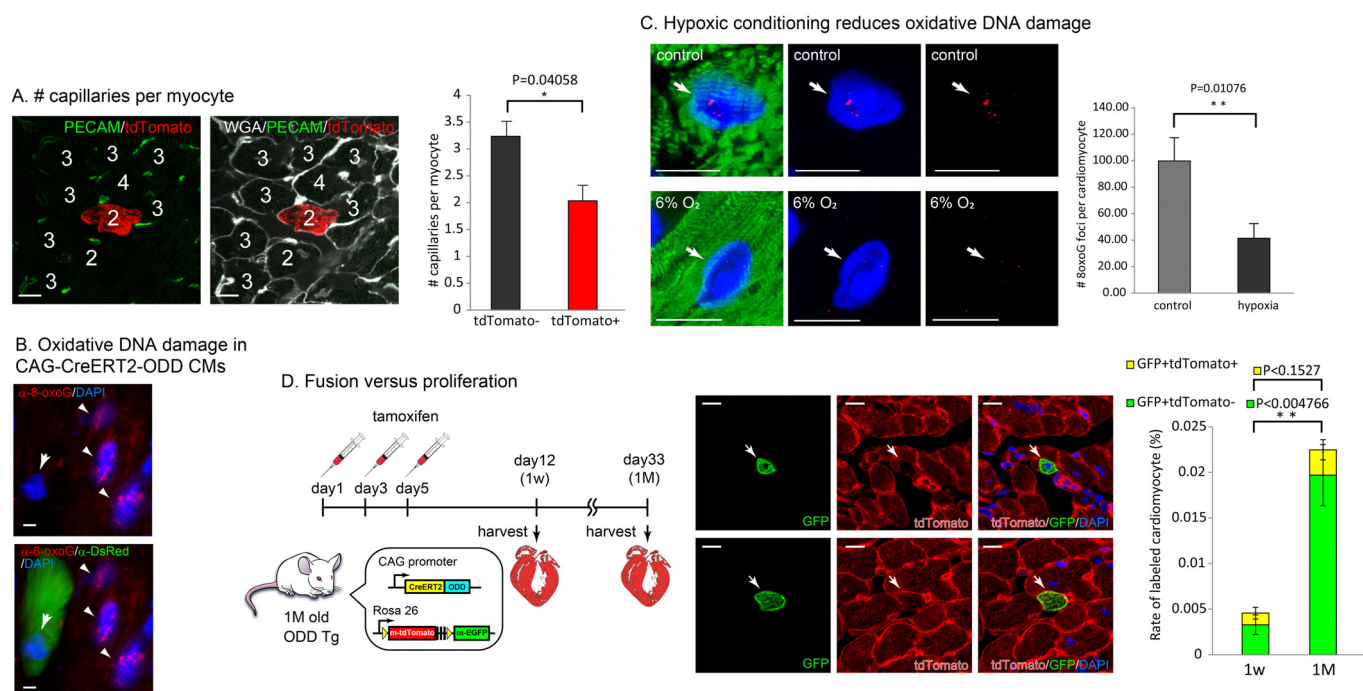
Extended Data Figure 3 | Hypoxic cardiomyocytes share characteristics of proliferative cardiomyocytes. **a**, tdTomato$^+$ cardiomyocytes in CAG-CreERT2–ODD;R26R/tdTomato mice at 1 week after tamoxifen administration were juxtaposed to fewer capillary blood vessels compared with surrounding non-labelled cardiomyocytes. **b**, Oxidative DNA damage indicated by immunostaining with an anti-8-oxoG antibody was lower in tdTomato$^+$ hypoxic cardiomyocytes compared with surrounding non-labelled cardiomyocytes at 1 week after tamoxifen administration. **c**, Immunostaining of cryosections with anti-8-oxoG (red) and anti-α-actinin (green) antibodies and quantification of nuclear foci in cardiomyocytes showed significantly decreased oxidative DNA damage after single exposure to 6% O_2 for 6 h. Scale bars indicate 10 μm. **d**, Cell fusion does not contribute mainly to the increase in number of tdTomato$^+$ cardiomyocytes. Schematic diagram shows the experiment using CAG-CreERT2–ODD;R26R/mTmG reporter mice. The number of eGFP$^+$/tdTomato$^-$ cardiomyocytes significantly increased at 1 month after tamoxifen pulse compared with 1 week after tamoxifen pulse, whereas the number of eGFP$^+$/tdTomato$^+$ cardiomyocytes did not. Data are presented as mean ± s.e.m. *$P < 0.05$, **$P < 0.01$. A two-tailed unpaired *t*-test was used for statistical analysis.
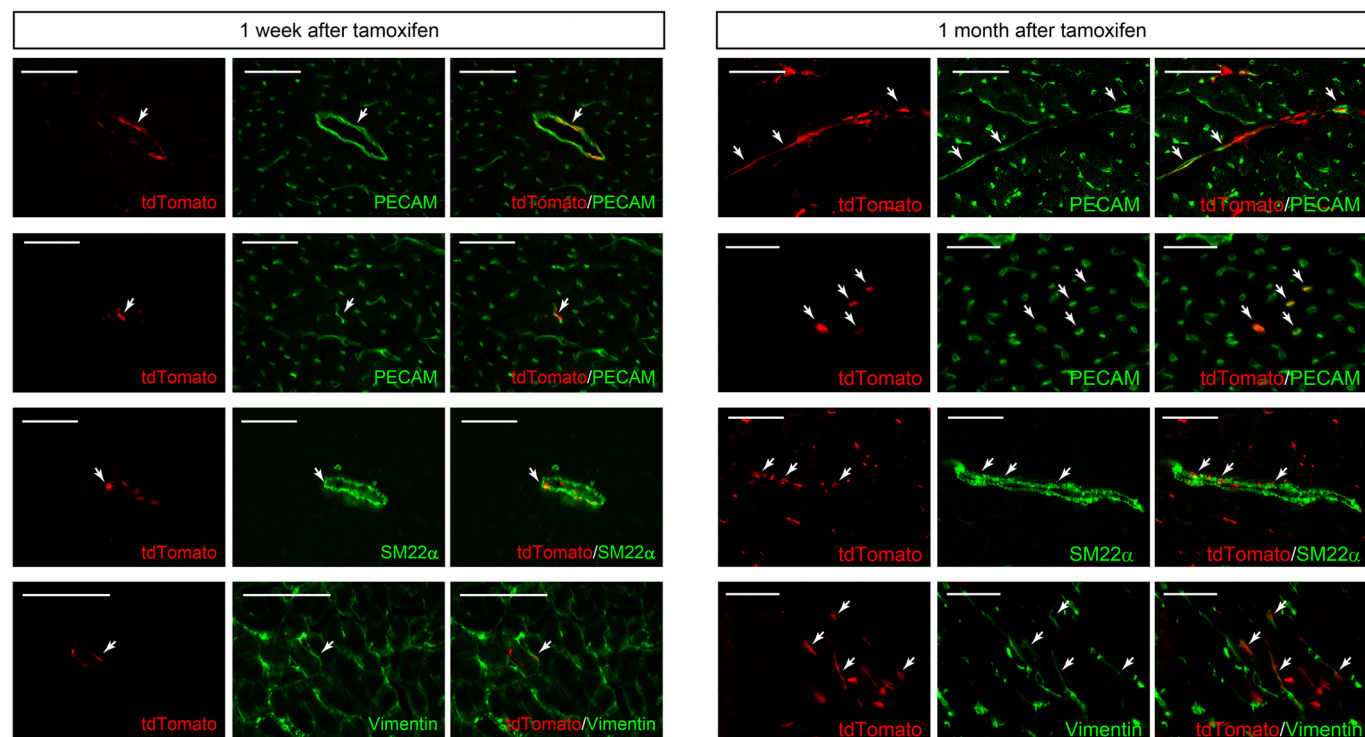
Extended Data Figure 4 | Hypoxic cells of non-cardiomyocyte lineage at 1 week and 1 month after tamoxifen administration to CAG-creERT2-ODD;R26R/tdTomato trandgenic mice. tdTomato$^+$ stabilized-Hif-1α cells were detected in endothelia in large vessels and capillaries (stained with anti-PECAM), vascular smooth muscle (stained with anti-SM22α) and interstitial fibroblasts (stained with anti-vimentin). Scale bars indicate 50 μm.
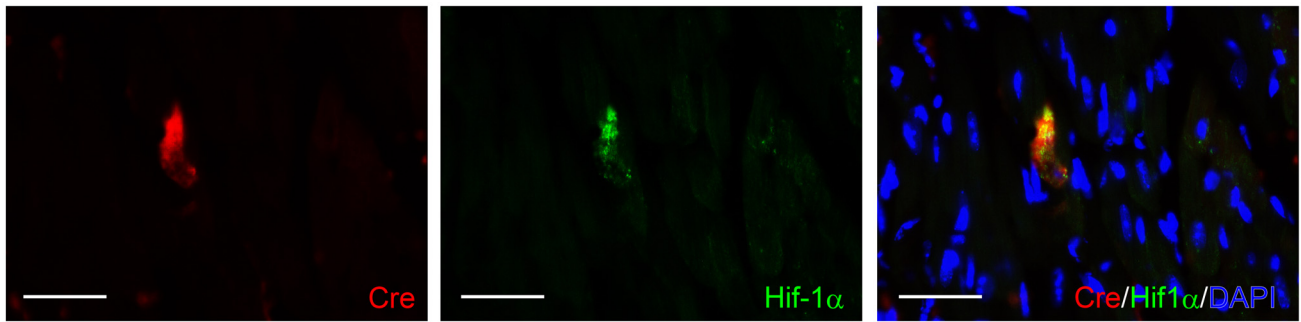
Extended Data Figure 5 | Co-immunostaining with anti-Cre and anti-Hif-1α antibodies showed co-localization of these two proteins in *αMHC-creERT2-ODD;R26R/tdTomato* transgenic hearts. Scale bars indicate 50 μm.

A. Laser microdissection
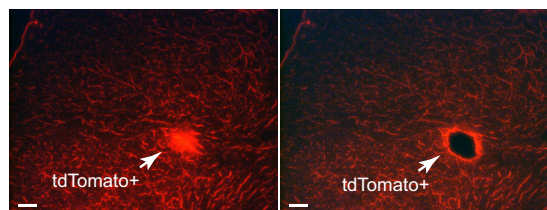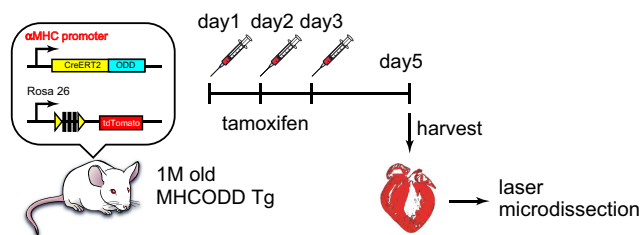
B. Hif-1α negative regulators

Gene name	Fold change	Description
Trp53	-6.1783	protein degradation
Os9	-2.2436	enhance oxygen dependent degradation
Naa10	-7.8829	protein degradation
Egln3	-7.7772	HIF-prolyl hydroxylase 3
Egln2	-6.6760	HIF-prolyl hydroxylase 2

C. Hif-1α positive regulators

Gene name	Fold change	Description
Usp20	3.1497	deubiquitination
Mtor	2.0784	translational activation

D. Hif-1α target genes

Gene name	Fold change	Description
Cyp4b1	9.2893	eicosanoid synthesis
Id2	8.9334	transcription factor
Bhlhe40	8.2701	transcription factor
Tfrc	8.0346	transferrin receptor
Nos3	5.2450	nitric oxide synthase
Cxcr4	4.4562	chemokine
Ctgf	4.0871	connective tissue growth factor
Pfkp	4.0568	phosphofructokinase
Pfkfb4	3.6031	phosphofructokinase
Car9	3.6031	carbonic anhydrase
Lep	2.8997	leptin
Hk2	1.7827	hexokinase
Mt1	1.4398	matrix metalloproteinase
Eno1	1.3458	enolase

E. Cdk/cyclins

Gene name	Fold change
Cdk5	9.5090
Cdk10	9.4275
Ccnc	4.2218
Ccnb1	4.0034
Cdk7	3.4526
Cdk9	2.0731
Cdk14	1.7773
Ccnd1	1.4764

F. negative cell cycle regulators

Gene name	Fold change
Cdkn1a	-8.5095
Cdkn1c	-8.0977
Mdm2	-7.5812
Trp53	-6.1783
Chek2	-5.8458
Brca2	-5.2007
Cdkn1b	-3.3223
Chek1	-2.6114

G. Hif genes

Gene name	Fold change
Hif1a	7.1792
Hif2a (Epas1)	0.4104
Hif1b (Arnt)	5.5112

H. Meis family transcription factors

Gene name	Fold change
Meis1	-4.1494
Meis2	-5.5345
Meis3	-7.0582

I. hypertrophy related genes

Gene name	Fold change
Hdac2	-7.8650
Nfatc4	-6.3181
Nppb	-5.4258
Hdac1	-3.9057
Hdac3	-2.5204

Extended Data Figure 6 | RNA-seq analysis of Hif-1α-related signalling pathway and cardiomyocyte cell cycle. a, Laser microdissection of tdTomato[+] cardiomyocytes. Hearts of αMHC-CreERT2–ODD;R26R/tdTomato mice were harvested 3 days after three doses of tamoxifen for three consecutive days. tdTomato[+] cardiomyocytes were collected from fresh cryosections. Scale bars indicate 10 µm. **b**, A table showing negative regulators of Hif-1α which are downregulated in tdTomato[+] cardiomyocytes. The $\log_2$ fold change of gene expression in tdTomato[+] cardiomyocytes compared with tdTomato[−] cardiomyocytes is shown. **c**, Positive regulators of Hif-1α which are upregulated in tdTomato[+] cardiomyocytes. **d**, A table showing known Hif-1α target genes upregulated in tdTomato[+] cardiomyocytes. **e**, Cyclin/CDKs upregulated in tdTomato[+] cardiomyocytes. A known cardiomyocyte cell cycle regulator cyclin D1 (Ccnd1) was consistently upregulated in tdTomato[+] stabilized-Hif-1α cardiomyocytes. **f**, Negative regulators of cell cycle including p21 (Cdkn1a), p57 (Cdkn1c), and DNA damage response related factors such as p53 (Trp53), Chek1, Chek2 and Brca2. **g**, Fold changes in Hif α and β subunits. **h**, All Meis family genes were downregulated in tdTomato[+] cardiomyocytes. **i**, A list of hypertrophy-related genes which are downregulated in tdTomato[+] cardiomyocytes.
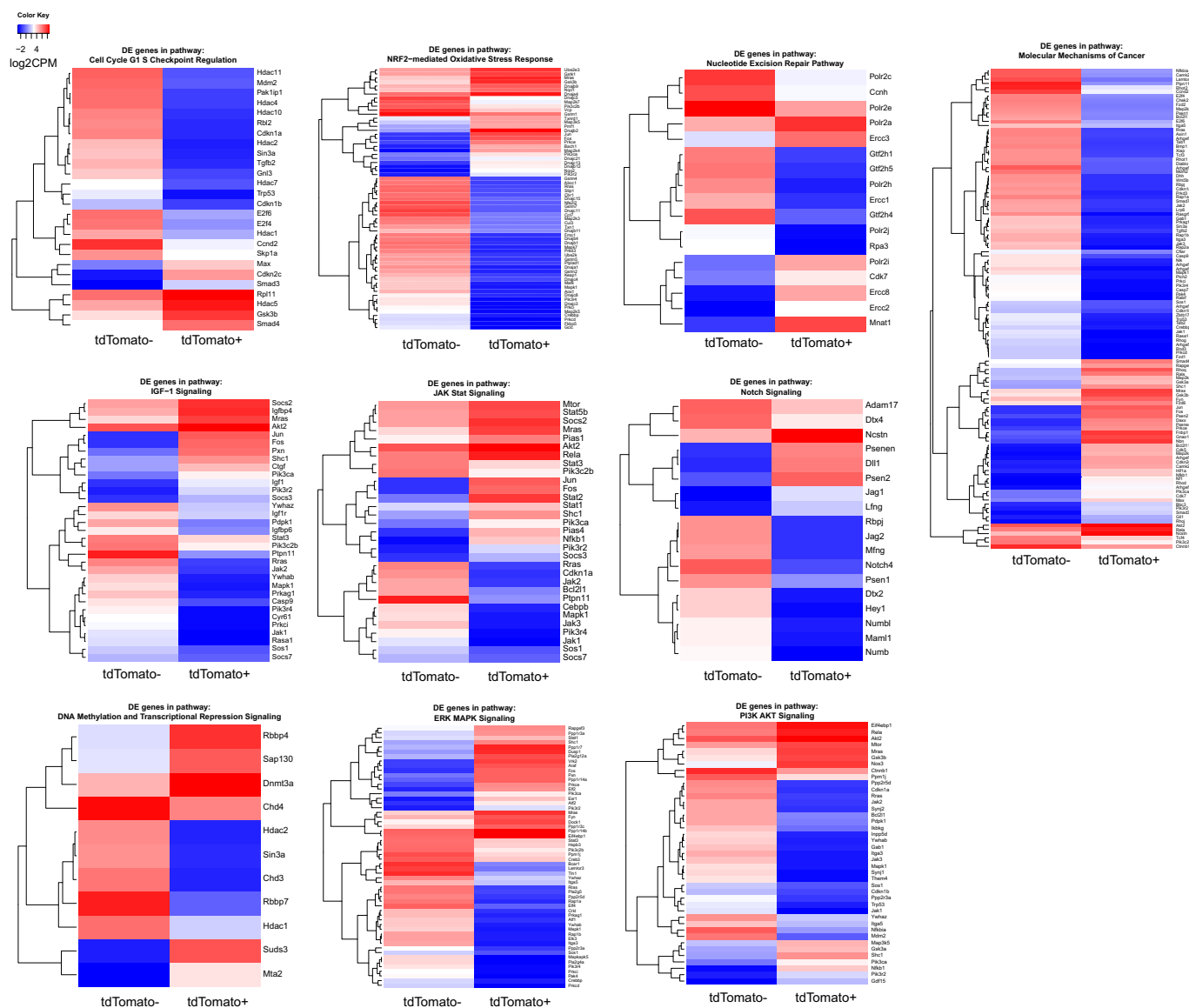
Extended Data Figure 7 | Selected heat maps showing differentially expressed genes involved in the ten most significantly altered pathways based on ingenuity pathway analysis.

A. Clonal expansion of tdTomato+ cardiomyocytes

Single

Paired

>2

- Single
- Paired
- >2

11.03±5.771%

26.41±5.322%

62.57±0.5627%

B. Mitosis

tdTomato

Ki67

tdTomato/Ki67/DAPI

p=0.02040

*

Fold difference in number of Ki67+ cardiomyocyte per field

tdTom-   tdTom+

C. MI

BrdU

BrdU/α-DsRed

BrdU/α-DsRed/DAPI

P=0.03792

*

Rate of tdTomato+ myocytes

Sham   MI

P=0.04376

*

Fold difference in number of BrdU+ cardiomyocyte per field

tdTom-   tdTom+

Extended Data Figure 8 | Hypoxic cardiomyocytes undergo cell cycle progression during normal cardiomyocyte turnover and in response to injury. **a**, WGA co-staining showed clusters of tdTomato[+] cardiomyocytes 1 month after tamoxifen pulse, indicating clonal expansion of cardiomyocytes. Scale bars indicate 10 μm. **b**, A marker for cycling cells, Ki67, was detected significantly more frequently in the nuclei of tdTomato[+] cardiomyocytes compared with that of tdTomato[−] cardiomyocytes. Scale bars indicate 10 μm. **c**, Acute myocardial infarction induces proliferation of hypoxic cardiomyocytes. Both the number of tdTomato[+] cardiomyocytes and BrdU[+] tdTomato[+] cardiomyocytes were increased within 1 week after coronary occlusion. Scale bars indicate 50 μm. *$P < 0.05$, **$P < 0.01$. A two-tailed unpaired t-test was used for statistical analysis.
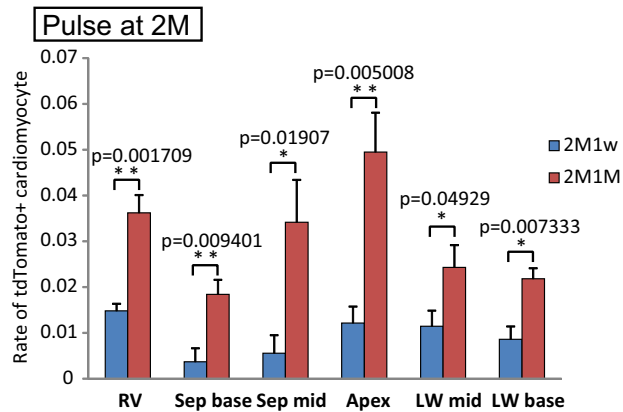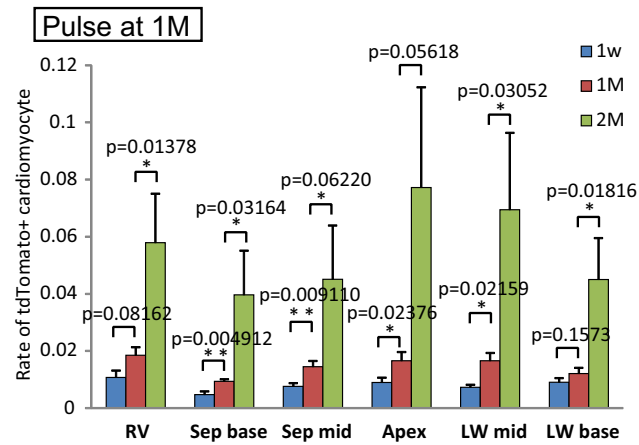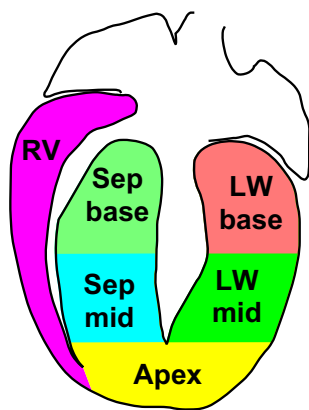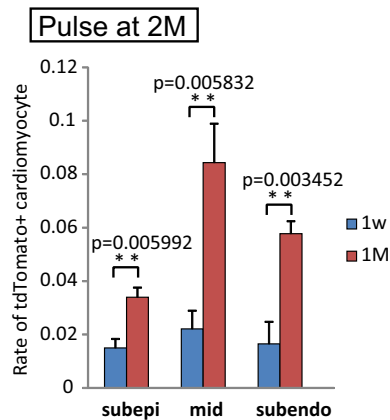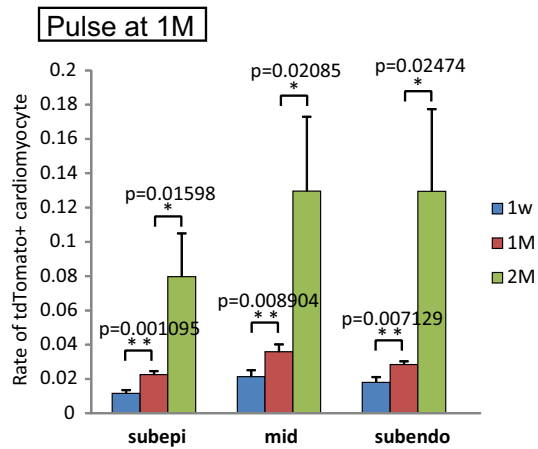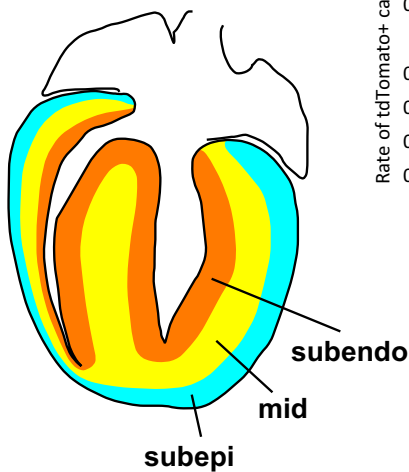
Extended Data Figure 9 | Distribution of hypoxic cardiomyocytes in the heart. **a**, The intra-cardiac localization of tdTomato$^+$ cardiomyocytes is depicted at six different levels: right ventricular lateral wall (RV), ventricular apex (Apex, a third of the ventricular myocardium from the apex of the heart), septum (Sep), and left ventricular lateral wall (LW); the septum and lateral wall are further divided into base (a third of the ventricular myocardium from the base of the heart) and middle (the region between base and apex. The graphs show the distribution of tdTomato$^+$ cardiomyocytes according to this classification. Upper graph shows the results from transgenic mice administered tamoxifen at 1 month of age, and lower graph shows the results from transgenic mice pulsed at 2 months of age. **b**, The localization of tdTomato$^+$ cardiomyocytes is depicted at three levels: subendo (a third of ventricular myocardium from endocardium), subepi (a third of ventricular myocardium from epicardium) and middle (the region between subendo and subepi). The graphs show the distribution of tdTomato$^+$ cardiomyocytes according to this classification. Upper graph shows the results from transgenic mice administered tamoxifen at 1 month of age, and lower graph shows the results from transgenic mice pulsed at 2 months of age. $*P < 0.05$, $**P < 0.01$. A two-tailed unpaired t-test was used for statistical analysis.
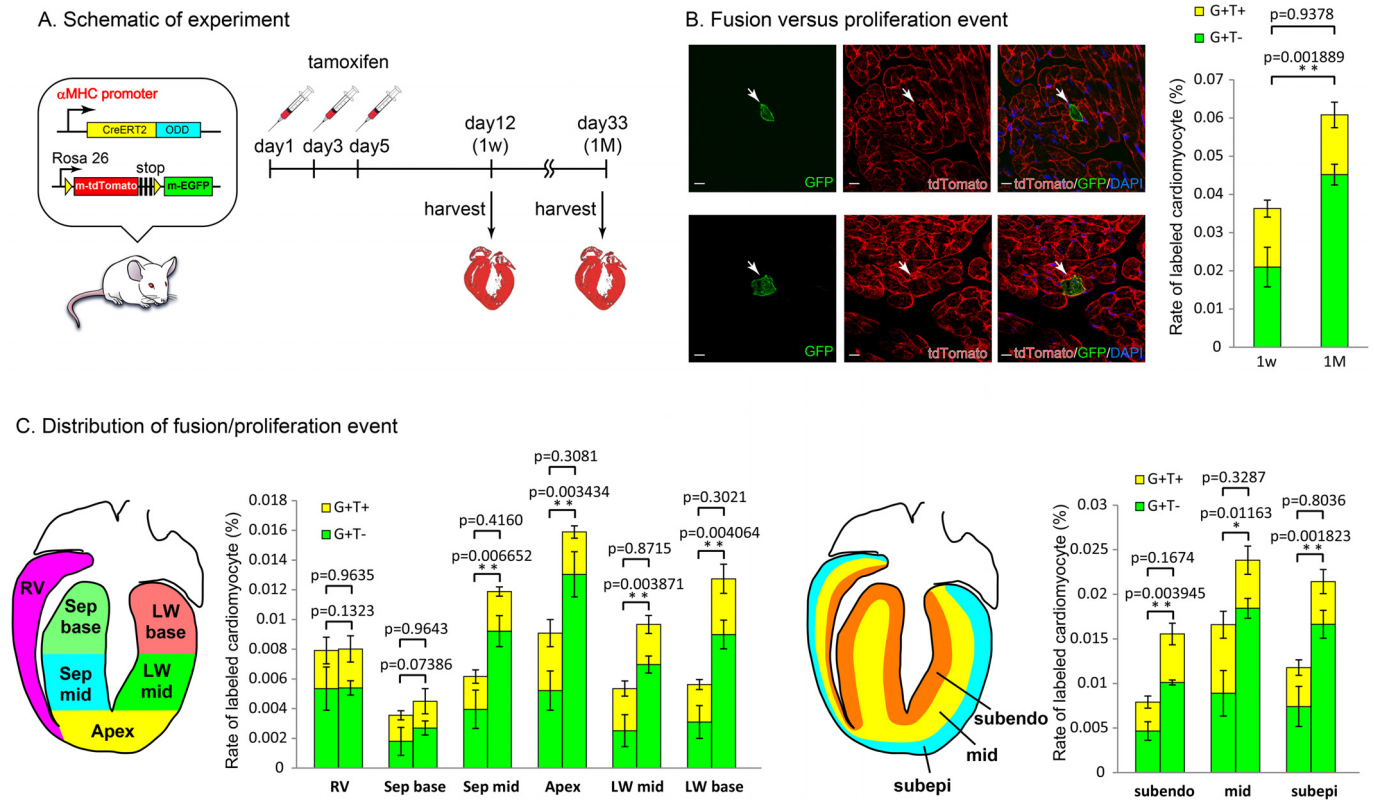
Extended Data Figure 10 | Cell fusion does not contribute significantly to the increase in number of hypoxic cardiomyocytes. a, Schematic diagram shows the experiment using αMHC-CreERT2–ODD;R26R/mTmG reporter mice. **b**, The number of eGFP$^+$/tdTomato$^-$ cardiomyocytes significantly increased at 1 month after tamoxifen pulse compared with the 1 week after tamoxifen pulse, whereas the number eGFP$^+$/tdTomato$^+$ cardiomyocytes did not. *$P < 0.05$, **$P < 0.01$. A two-tailed unpaired t-test was used for statistical analysis.

Melanoma–intrinsic β–catenin signalling prevents anti–tumour immunity

Stefani Spranger[1], Riyue Bao[2] & Thomas F. Gajewski[1,3]

Melanoma treatment is being revolutionized by the development of effective immunotherapeutic approaches[1,2]. These strategies include blockade of immune-inhibitory receptors on activated T cells; for example, using monoclonal antibodies against CTLA-4, PD-1, and PD-L1 (refs 3–5). However, only a subset of patients responds to these treatments, and data suggest that therapeutic benefit is preferentially achieved in patients with a pre-existing T-cell response against their tumour, as evidenced by a baseline CD8$^+$ T-cell infiltration within the tumour microenvironment[6,7]. Understanding the molecular mechanisms that underlie the presence or absence of a spontaneous anti-tumour T-cell response in subsets of cases, therefore, should enable the development of therapeutic solutions for patients lacking a T-cell infiltrate. Here we identify a melanoma-cell-intrinsic oncogenic pathway that contributes to a lack of T-cell infiltration in melanoma. Molecular analysis of human metastatic melanoma samples revealed a correlation between activation of the WNT/β-catenin signalling pathway and absence of a T-cell gene expression signature. Using autochthonous mouse melanoma models[8,9] we identified the mechanism by which tumour-intrinsic active β-catenin signalling results in T-cell exclusion and resistance to anti-PD-L1/anti-CTLA-4 monoclonal antibody therapy. Specific oncogenic signals, therefore, can mediate cancer immune evasion and resistance to immunotherapies, pointing to new candidate targets for immune potentiation.

To identify oncogenic pathways inversely associated with T-cell infiltration, we categorized 266 metastatic human cutaneous melanoma samples into those with low (non-T-cell-inflamed) and high (T-cell-inflamed) expression of T-cell signature genes[6,10] (Fig. 1a). Comparative gene expression profiling revealed 1,755 genes that were preferentially expressed in the non-inflamed patient cohort ($q < 0.01$) (Supplementary Table 1). Pathway analysis, comparing 91 non-T-cell-inflamed to 106 T-cell-inflamed patients, indicated active β-catenin signalling (*APC2*, *SOX2*, *SOX11* and *WNT7B*; $P = 0.00116$) as well as dermatan-sulfate biosynthesis (*HS6ST2* and *NDST3*; $P = 0.00196$) in the non-T-cell-inflamed cohort. Previous reports suggested that active β-catenin signalling in melanoma was associated with more aggressive disease[9]. To determine if activation of the β-catenin pathway might be modified by specific mutations, we analysed exome-sequencing data for all 197 patients. Indeed, seven tumour samples (7.7%) with the non-T-cell-inflamed phenotype showed gain-of-function mutations in β-catenin (*CTNNB1*), versus one case in the T-cell-infiltrated cohort. Additionally, loss-of-function mutations in negative regulators of the pathway (*APC*, *AXIN1*, *TCF1*) were identified in ten non-T-cell-inflamed tumours (11%) (Supplementary Table 3). To identify the total percentage of tumours with an active β-catenin pathway, we assessed expression of six well-characterized β-catenin target genes[11]. Forty-eight per cent (44 patients) in the non-T-cell-inflamed subset showed expression of at least five of the six β-catenin target genes versus 3.8% (4 patients) of the T-cell-inflamed tumours (Fig. 1b). While several cases were associated with defined mutations (*CTNNB1*, 14%; *APC*, *AXIN1* or *TCF1*, 23%) the majority (61%) of the remaining cases

showed increased expression of either WNT7B (WNT7B, 29.5%; 13 patients), FZD3 (FZD3, 20.5%; 9 patients), or β-catenin itself (11%; 5 patients; Supplementary Table 3). In sum, an increased *CTNNB1* score was predictive for the lack of T cells, with an odds ratio of 4.9 (Extended Data Fig. 1a). Additional analysis revealed a negative correlation between individual β-catenin target genes and *CD8A* transcripts, which was opposite to the pattern of PD-L1 expression (Fig. 1c and Supplementary Table 2)[12]. Immunohistochemical analysis of an independent sample cohort also revealed an inverse association between stabilized β-catenin and CD8$^+$ T cells (Fig. 1d and Extended Data Fig. 1b).

We investigated directly whether active β-catenin signalling within tumour cells could adversely affect anti-tumour T-cell responses using inducible autochthonous mouse models (genetically engineered mice (GEM)) driven by conditional active Braf with or without conditional PTEN deletion and expression of active β-catenin. These GEMs developed tumours with similar latency, as reported previously (Fig. 1e and Extended Data Fig. 2a–c)[8,9]. We focused on *Braf*^{V600E}/*Pten*$^{-/-}$ and *Braf*^{V600E}/*Pten*$^{-/-}$/*CAT-STA* mice due to the similar rate of onset of tumour development in these strains (Extended Data Fig. 2b, c). Using gene array analysis and histological examination we confirmed that the developing tumours were indeed melanomas (Extended Data Fig. 2d, e)[8,9], albeit with less pigmentation in *Braf*^{V600E}/*Pten*$^{-/-}$ tumours (Extended Data Figs 2e, f and 3a, b)[9]. Analysis of immune infiltrates revealed that *Braf*^{V600E}/*Pten*$^{-/-}$ tumours indeed contained CD3$^+$ T cells. However, tumours with active β-catenin showed almost a complete absence of T cells (Fig. 1f). Fluorescent immunohistology confirmed the absence of intra-tumoural CD3$^+$ T cells in *Braf*^{V600E}/*Pten*$^{-/-}$/*CAT-STA* tumours (Fig. 1g and Extended Data Fig. 3a–c) with only rare T cells observed in the epidermis. These results indicate that tumour-intrinsic β-catenin activation dominantly excludes T-cell infiltration into the melanoma tumour microenvironment.

The T-cell infiltrate in *Braf*^{V600E}/*Pten*$^{-/-}$ tumours consisted of both CD4$^+$ and CD8$^+$ T cells, with the majority of them expressing the αβ-T-cell antigen receptor (TCR) (Extended Data Fig. 4a, b). The majority were CD44^{hi}/CD62L^{lo}/CD45RA^{lo}, suggesting an activated phenotype (Extended Data Fig. 4c), and 6% FoxP3$^+$ regulatory T cells were detected (Extended Data Fig. 4d). Additionally, CD8$^+$ T cells from *Braf*^{V600E}/*Pten*$^{-/-}$ tumours showed expression of PD-1 and Lag3 (Extended Data Fig. 4e, f), markers of T-cell dysfunction in the tumour context[13]. Consistent with this phenotype, sorted CD3$^+$ T cells from *Braf*^{V600E}/*Pten*$^{-/-}$ tumours showed defective interleukin (IL)-2 production but were capable of producing interferon (IFN)-γ (Extended Data Fig. 4g, h). Comparable studies on the few T cells from *Braf*^{V600E}/*Pten*$^{-/-}$/*CAT-STA* tumours showed predominantly a naive phenotype (Extended Data Fig. 4a–e). Correspondingly, increased PD-L1 expression in *Braf*^{V600E}/*Pten*$^{-/-}$ tumours was observed, consistent with previous work linking PD-L1 expression with the presence of CD8$^+$ T cells (Extended Data Fig. 4i, j)[12]. We did not detect significant differences in CD11b$^+$Gr1$^+$ myeloid-derived suppressor cells
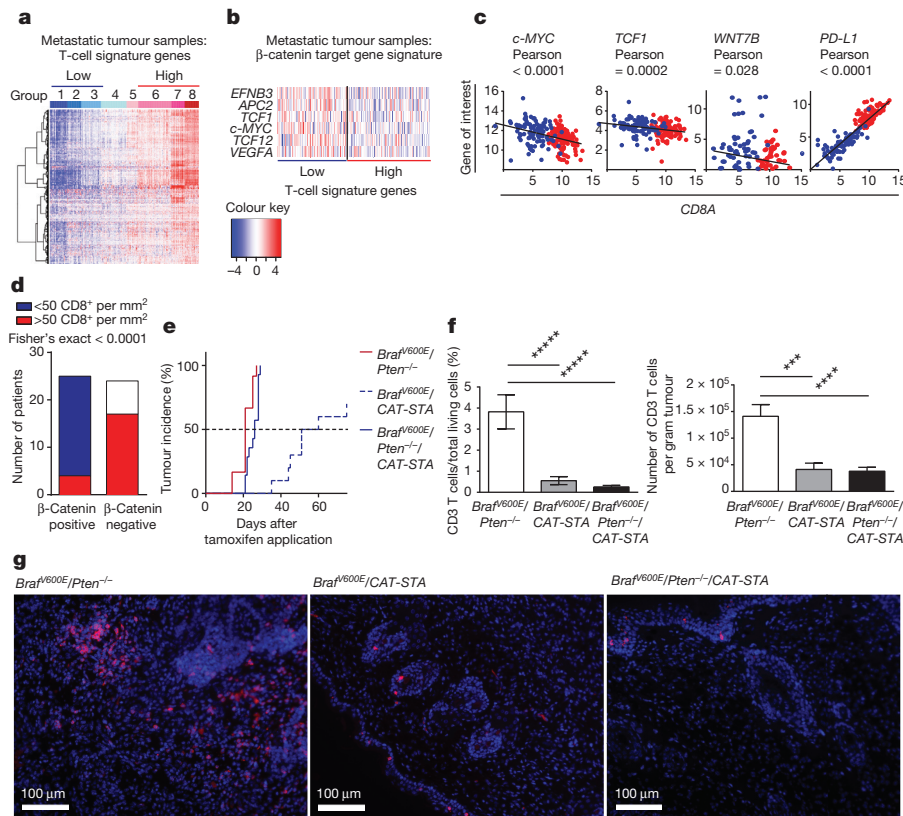
Figure 1 | Melanoma-intrinsic β-catenin pathway activation correlates with T-cell exclusion. a, b, Heat maps of 266 metastatic melanomas clustered in low versus high T-cell signature gene groups (**a**), and β-catenin target genes within the T-cell-signature high and low cohorts (**b**). **c,** Pearson correlation of *CD8A* expression with c-MYC, TCF1 and WNT7B (red indicates T-cell-signature high, blue indicates T-cell-signature low). **d,** Correlation between β-catenin and CD8 in melanoma biopsies. Fisher's exact test with $n = 49$. **e,** Tumour incidence rates of GEMs (median time to tumour event): $Braf^{V600E}/Pten^{-/-}$:

100%, 21 days ($n = 14$); $Braf^{V600E}/CAT\text{-}STA$: 85%, 55.5 days ($n = 8$); $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$: 100%, 26 days ($n = 14$). **f,** CD3$^+$ T cells depicted as percentage living cells and absolute numbers per gram tumour. $n = 20$, mean ± standard error of the mean (s.e.m.), Mann–Whitney U test. **g,** Representative example out of five for fluorescent immunohistochemistry staining against CD3$^+$ T cells. Scale bars, 100 μm. See Extended Data Fig. 3 for overview. ***$P \leq 0.001$, ****$P \leq 0.0001$, *****$P \leq 0.00001$.

($Braf^{V600E}/Pten^{-/-}$, 1,047 ± 418, to $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$, 739 ± 185 cells per gram tumour; $P = 0.7429$) (Extended Data Fig. 4k)[14].

Although the models used in this study recapitulate defined carcinogenic processes, one drawback is the potentially low number of generated neo-antigens, which may lead to reduced immunogenicity[15]. To circumvent this we crossed both GEMs to a mouse strain allowing Cre-dependent expressing of the model antigen SIYRYYGL (SIY)[16]. We investigated whether lack of T-cell infiltration into the $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours was secondary to a lack of initial T-cell priming by adoptive transfer of carboxyfluorescein succinimidyl ester (CFSE)-labelled SIY-specific TCR-transgenic 2C T cells. While SIY-negative mice failed to accumulate 2C T cells within the tumour-draining lymph nodes (TdLNs) or the tumour site, SIY-positive mice had detectable 2C T cells in the TdLNs in both GEMs. However, no proliferation of 2C T cells was identified within the TdLNs in the $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA/SIY^+$ model, whereas activation of T cells within the TdLNs of $Braf^{V600E}/Pten^{-/-}$ mice was brisk (Fig. 2a, b). Accordingly, the presence of proliferated 2C T cells was observed at the tumour site exclusively in $Braf^{V600E}/Pten^{-/-}$ mice (Fig. 2a, b). These data indicate that tumour-intrinsic β-catenin signalling prevents the early steps of T-cell priming against tumour-associated antigens.

The absence of early T-cell priming in $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumour-bearing mice suggested a defect in the antigen-presenting-cell compartment. Work using transplantable tumour models has indicated that Batf3-lineage dendritic cells are crucial for cross-presentation of tumour antigens to CD8$^+$ T cells[17–19]. Dendritic cell subsets

(CD45$^+$MHCII$^+$CD11c$^+$) were analysed phenotypically within the tumour microenvironment with minimal differences observed in the number of conventional dendritic cells (B220$^-$), plasmacytoid dendritic cells (B220$^+$), monocytes (B220$^-$Ly6C$^+$), or Langerhans dendritic cells (B220$^-$CD207$^+$). Strikingly, the CD8α$^+$ and CD103$^+$ dendritic cell populations were nearly completely absent from $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours (Fig. 2c–e). CD103$^+$ dendritic cells were also reduced in the TdLNs, while being preserved in the spleen (Fig. 2d, e; data not shown). Sorted tumour-infiltrating CD45$^+$CD11c$^+$ dendritic cells from $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours also showed reduced expression of the CD103$^+$ dendritic cell transcripts Batf3, Irf8 and Itgae (Extended Data Fig. 5a, b), and dendritic cells showed reduced expression of the key innate cytokine IFN-β (Extended Data Fig. 5a). Together, these results suggest that the failed T-cell priming against tumour-associated antigen in $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours is secondary to defective recruitment and activation of Batf3-lineage dendritic cells.

To determine whether T-cell infiltration into $Braf^{V600E}/Pten^{-/-}$ tumours was dependent on CD103$^+$ dendritic cells, $Batf3^{-/-}$ bone marrow chimaeras were generated. Indeed, tumours from $Braf^{V600E}/Pten^{-/-}/Batf3^{-/-}$ bone marrow chimaeras failed to develop T-cell infiltration (Fig. 2f). To assess whether poor dendritic cell recruitment was indeed the major functional barrier, we generated Flt3 ligand-derived bone-marrow dendritic cells activated with polyinosinic:polycytidylic acid (poly(I:C))[20] for intra-tumoural injection, which were found to restore T-cell infiltration in $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours (Fig. 2h) and led to a modest reduction in tumour weight
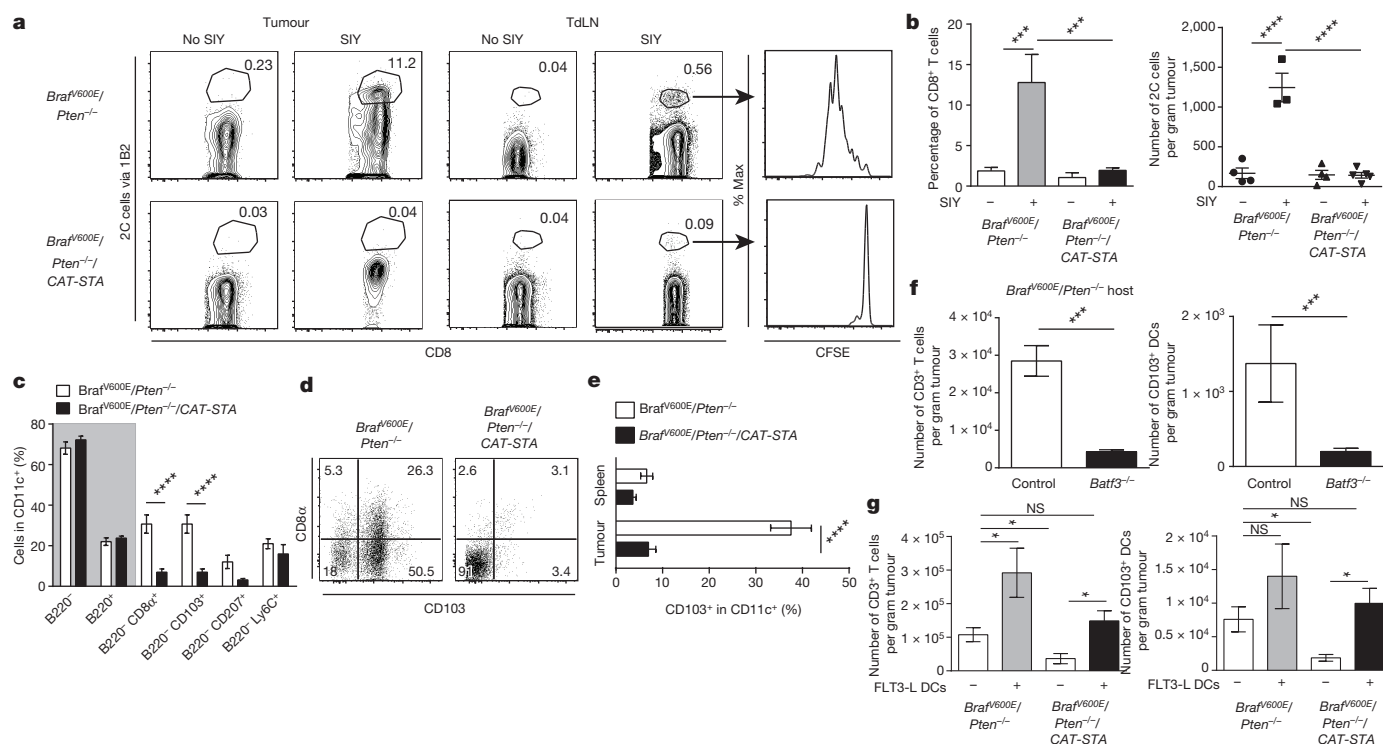
Figure 2 | Braf^V600E^/Pten^−/−^/CAT-STA mice show impaired priming of anti-tumour T cells and reduced numbers of CD103^+^ dermal dendritic cells. **a**, Abundance and proliferation of TCR-transgenic 2C T cells. Depicted are representative examples pre-gated on live, CD45^+^CD3^+^CD8^+^ cells. **b**, Statistical analysis of **a** ($n = 8$). **c**, Percentages of dendritic cell subsets within Braf^V600E^/Pten^−/−^ and Braf^V600E^/Pten^−/−^/CAT-STA tumours ($n = 8$). **d**, Representative example of CD103/CD8α staining (gated CD45^+^MHCII^hi^CD11c^+^).

e, Quantification of CD103^+^ dendritic cells ($n = 12$). **f**, Amount of CD3^+^ T cell and CD103^+^ dendritic cell (DC) infiltration in Braf^V600E^/Pten^−/−^ tumours reconstituted with control or Batf3^−/−^ bone marrow ($n = 4$ and $n = 11$, respectively). **g**, Intra-tumoural injection of Flt3 ligand-derived dendritic cells into Braf^V600E^/Pten^−/−^/CAT-STA tumours ($n = 6$ control mice and 8 mice, PBS control). All data are mean ± s.e.m., Mann–Whitney U test. $*P \leq 0.05$, $***P \leq 0.001$, $****P \leq 0.0001$; NS, not significant.

(Extended Data Fig. 5c). Using dendritic cells generated from actin–green fluorescent protein (GFP) transgenic mice, injected dendritic cells were retained within the tumour microenvironment during this experimental timeframe (Extended Data Fig. 5d). Together, these results suggest that the major immunological defect in the context of melanomas expressing tumour-intrinsic β-catenin signalling is defective recruitment of CD103^+^ dendritic cells.

To pursue mechanisms explaining failed CD103^+^ dermal dendritic cell recruitment, gene expression profiling was performed from tumours of the two genotypes, focusing on chemokines (Supplementary Table 4). Five chemokines were differentially expressed, with four of these (CCL3, CXCL1, CXCL2 and CCL4) being expressed at lower levels in Braf^V600E^/Pten^−/−^/CAT-STA tumours (Fig. 3a, b and Supplementary Table 4). For evaluation of tumour-cell-intrinsic chemokine production *in vivo*, we crossed Braf^V600E^/Pten^−/−^ mice to yellow fluorescent protein (YFP)-reporter mice, which allowed identification of transformed YFP^+^ cells. Ccl4 transcripts were detected exclusively in the YPF^+^ cell population from Braf^V600E^/Pten^−/−^ mice, while control sorted YFP^+^ cells from Braf^WT^/Pten^−/−^ mice or YFP^−^ cells showed no detectable Ccl4 (Fig. 3c). A similar expression pattern was observed for CXCL1, whereas CCL3 and CXCL2 were expressed by normal melanocytes and stromal cells, respectively (Fig. 3c). CD45^+^CD3^+^ and CD45^+^CD3^−^ cells sorted as controls, showing the expected patterns of Ifng and Ifnb expression (Extended Data Fig. 6c, d). Expression analysis of the corresponding chemokine receptor, Ccr5, revealed a lack of CCR5 expression by the dendritic cells isolated from Braf^V600E^/Pten^−/−^/CAT-STA tumours (Fig. 3d). CCR5 has previously been linked with the migratory capacity of CD8α^+^ dendritic cells[21]. To confirm this observation, we generated tumour cell lines from both GEMs and found increased production of CCL4 by BP (Braf^V600E^/Pten^−/−^-derived) tumour cells compared to

BPC (Braf^V600E^/Pten^−/−^/CAT-STA-derived) tumour cells (Extended Data Fig. 6a, b). To strengthen a functional role for CCL4, we used an *in vitro* migration assay in response to recombinant murine CCL4 as well as tumour cell line supernatants (Fig. 3e). Indeed, skin-derived CD11c^+^CD103^+^ dendritic cells and lymph-node-derived dendritic cells (CD11c^+^CD8α^+^) migrated in response to CCL4 and BP supernatants but not to BPC supernatants. Together, these results indicate that failed recruitment of CD103^+^ dendritic cells into the tumour microenvironment of Braf^V600E^/Pten^−/−^/CAT-STA tumours was, at least in part, due to defective production of the chemokine CCL4.

We then pursued a mechanism by which β-catenin activation might prevent Ccl4 gene expression, since CCL4 has also been associated with a T-cell infiltrate in human melanoma tumours[6,22]. Previous reports had suggested that Wnt/β-catenin signalling induces expression of the transcriptional repressor ATF3 (ref. 23), and that ATF3 suppresses Ccl4 (ref. 24). Indeed, Atf3 was expressed at higher levels in primary tumours as well as in BPC tumour cell lines from Braf^V600E^/Pten^−/−^/CAT-STA mice (Fig. 3f). A chromatin immunoprecipitation (ChIP) assay revealed binding of ATF3 to the Ccl4 promoter region in Braf^V600E^/Pten^−/−^/CAT-STA cells while no binding was observed for Ccl2, a chemokine lacking an ATF3-binding site (Fig. 3g). Short interfering RNA (siRNA)-mediated knockdown of Atf3 or Ctnnb1 in BPC tumour cells restored CCL4 production (Fig. 3h). To examine this relationship in human melanoma, we analysed two melanoma cell lines, mel537 and mel888, which show low or high β-catenin expression, respectively (Extended Data Fig. 7a, c). Consistent with the murine cell lines, increased ATF3 and decreased CCL4 production were observed in the β-catenin-positive mel888 cells (Extended Data Fig. 7b, e), and increased binding of ATF3 to the CCL4 promoter was also detected (Extended Data Fig. 7d). siRNA-mediated knockdown of ATF3 or
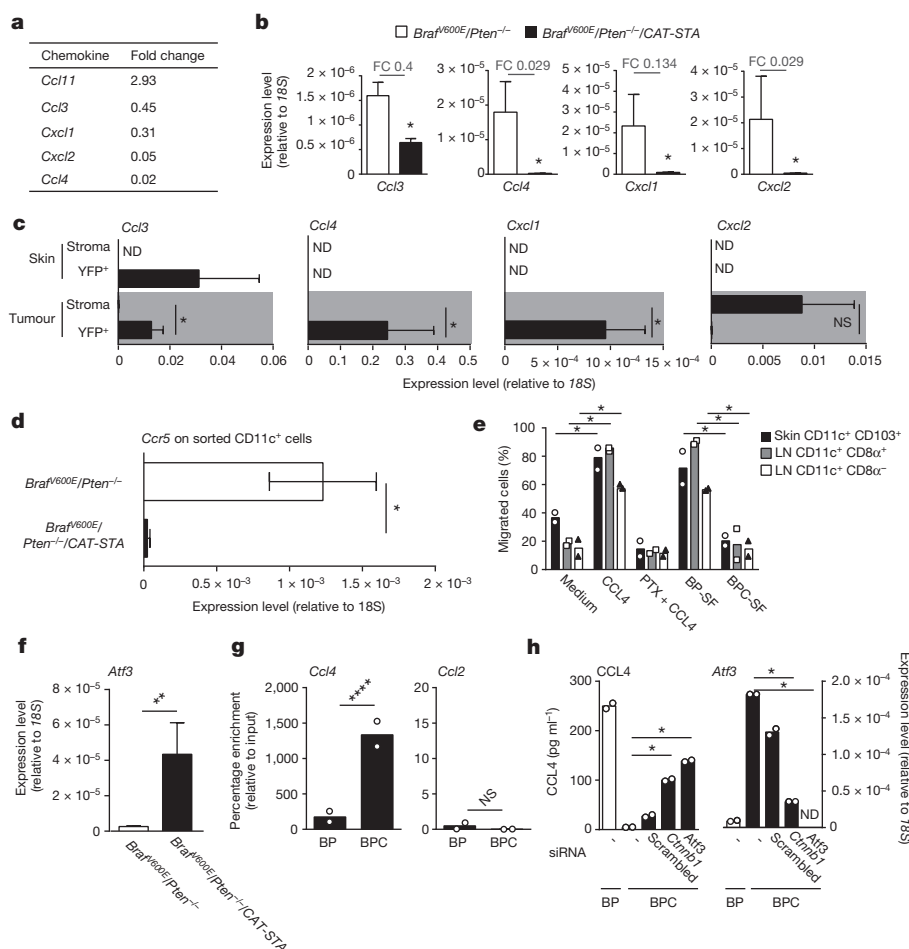
Figure 3 | Active β-catenin signalling within tumour cells suppresses the recruitment of CD103+ dendritic cells. a, Chemokine expression in GEM tumours assessed via gene array analysis ($n = 4$). **b,** Confirmatory quantitative polymerase chain reaction with reverse transcription (qRT–PCR) ($n = 8$) with fold change (FC) indicated at the top. **c,** Transcript levels of *Ccl3, Ccl4, Cxcl1* and *Cxcl2* assessed from YFP+ and CD45− YFP− cells from *Braf^V600E/Pten^−/−/* YFP+ tumours ($n = 5$), sorted on day 7 after tamoxifen administration. ND, not detected. **d,** Expression level of CCR5 in sorted CD45+CD11c+ dendritic cells ($n = 8$). **e,** Migration assay of dendritic cell subsets towards recombinant mouse CCL4 or conditioned medium (SF) (two independent experiments, duplicates per experiment). **f,** *Atf3* transcripts in tumour tissues ($n = 8$). **g,** ATF3-specific ChIP assay in BP and BPC cell lines (two independent experiments, duplicates per experiment). **h,** Amount of secreted CCL4 in 48-h-conditioned siRNA-treated tumour-cell BP and BPC supernatants, assessed by enzyme-linked immunosorbent assay (ELISA) and *Atf3* expression at the endpoint detected by qRT–PCR (two independent experiments, duplicates per experiment). All data are mean ± s.e.m., Mann–Whitney *U* test. *$P \leq 0.05$, **$P \leq 0.01$, ****$P \leq 0.0001$; NS, not significant.

β-catenin in mel888 cells restored CCL4 production (Extended Data Fig. 7e).

We additionally investigated whether decreased presence of BATF3-lineage dendritic cells was associated with active β-catenin signalling in human melanoma metastases. A Pearson correlation analysis for expression of *THBD* (CD141, marker for human BATF3-lineage dendritic cells[25]; $P < 0.0001$), *BATF3* ($P = 0.0336$) and *IRF8* ($P < 0.0001$) revealed a negative association with the *CTNNB1* score (Extended Data Fig. 8 and data not shown). Furthermore, *CCL4* had already been observed to correlate positively with T-cell transcripts (Fig. 1a). We conclude that β-catenin activation within melanoma cells results in decreased *CCL4* gene expression, which is at least partly mediated through ATF3-dependent transcriptional repression (Extended Data Fig. 9).

To explore the therapeutic relevance of the lack of T-cell infiltration, both GEMs were treated with a combination of anti-CTLA-4 and anti-PD-L1 monoclonal antibodies[3,26]. While treatment of *Braf^V600E/Pten^−/−* mice resulted in a significant delay in tumour outgrowth, no therapeutic effect was detected in *Braf^V600E/Pten^−/−/CAT-STA* mice (Fig. 4a, b). To evaluate whether restoration of intra-tumoural dendritic cells could restore immunotherapy responsiveness, Flt3 ligand-induced bone-marrow dendritic cells were injected intra-tumourally into *Braf^V600E/Pten^−/−/CAT-STA* tumours. Indeed, introduction of dendritic cells had a partial therapeutic effect, which was improved significantly with anti-CTLA-4 and anti-PD-L1 monoclonal antibodies (Fig. 4c).

We conclude that melanoma-cell-intrinsic activation of an oncogenic pathway can result in exclusion of the host immune response, including the absence of a T-cell infiltrate within the tumour microenvironment. Although 48% of non-T-cell-infiltrated melanomas show active β-catenin signalling, it is conceivable that additional oncogenic signalling pathways might mediate immune exclusion in other cases. The WNT/β-catenin pathway may contribute to immune evasion in other tumour entities beyond melanoma, which would be consistent with previous *in vitro* work[27]. Within T cells, β-catenin appears to inhibit T-cell activation, suggesting that a general immune-potentiating effect may result from therapeutic targeting[28]. The T-cell-inflamed tumour microenvironment phenotype appears to be predictive of clinical response to immune-based therapies[7,10,29]. Immune escape among this subset appears to be a consequence of dominant effects of negative regulatory pathways such as PD-1, arguing that the clinical activity of anti-PD-1 is tipping the balance in favour of an ongoing immune response[12]. By inference, tumour-intrinsic
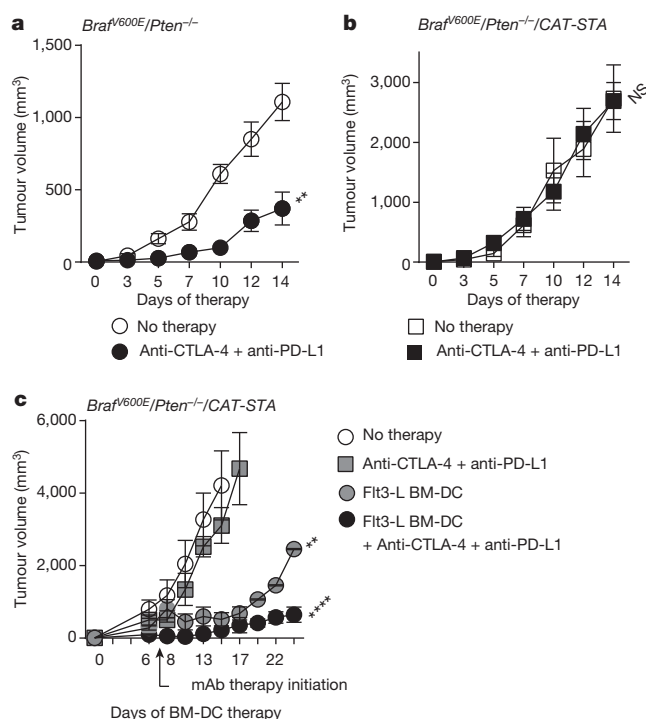
Figure 4 | Reconstitution with Flt3 ligand dendritic cells reverses resistance to immunotherapy. a, b, Tumour growth in $Braf^{V600E}/Pten^{-/-}$ (**a**) and $Braf^{V600E}/Pten^{-/-}/CAT$-$STA$ (**b**) mice untreated or treated with anti-CTLA-4 and anti-PD-L1 therapy ($n = 10$). **c,** Tumour growth of $Braf^{V600E}/Pten^{-/-}/CAT$-$STA$ tumour-bearing mice that were untreated, treated with anti-CTLA-4 and anti-PD-L1 therapy, intra-tumoural Flt3 ligand (Flt3-L) dendritic cell injections, or combination therapy ($n = 5$). BM-DC, bone-marrow dendritic cell; mAb, monoclonal antibody. All data are mean ± s.e.m., two-way analysis of variance (ANOVA) test. **$P \leq 0.01$, ***$P \leq 0.0001$; NS, not significant.

β-catenin activation may represent one mechanism of primary resistance to these therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Kaufman, H. L. et al. The Society for Immunotherapy of Cancer consensus statement on tumour immunotherapy for the treatment of cutaneous melanoma. Nature Rev. Clin. Oncol. **10,** 588–598 (2013).
2. Mellman, I., Coukos, G. & Dranoff, G. Cancer immunotherapy comes of age. Nature **480,** 480–489 (2011).
3. Wolchok, J. D. et al. Nivolumab plus ipilimumab in advanced melanoma. N. Engl. J. Med. **369,** 122–133 (2013).
4. Topalian, S. L. et al. Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. J. Clin. Oncol. **32,** 1020–1030 (2014).
5. Hodi, F. S. et al. Improved survival with ipilimumab in patients with metastatic melanoma. N. Engl. J. Med. **363,** 711–723 (2010).
6. Harlin, H. et al. Chemokine expression in melanoma metastases associated with CD8+ T-cell recruitment. Cancer Res. **69,** 3077–3085 (2009).
7. Ji, R. R. et al. An immune-active tumor microenvironment favors clinical response to ipilimumab. Cancer Immunol. Immunother. **61,** 1019–1031 (2012).
8. Dankort, D. et al. Braf(V600E) cooperates with Pten loss to induce metastatic melanoma. Nature Genet. **41,** 544–552 (2009).
9. Damsky, W. E. et al. β-Catenin signaling controls metastasis in Braf-activated Pten-deficient melanomas. Cancer Cell **20,** 741–754 (2011).
10. Galon, J. et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science **313,** 1960–1964 (2006).
11. Rimm, D. L., Caca, K., Hu, G., Harrison, F. B. & Fearon, E. R. Frequent nuclear/cytoplasmic localization of β-catenin without exon 3 mutations in malignant melanoma. Am. J. Pathol. **154,** 325–329 (1999).
12. Spranger, S. et al. Up-regulation of PD-L1, IDO, and T_{regs} in the melanoma tumor microenvironment is driven by CD8+ T cells. Sci. Transl. Med. **5,** 200ra116 (2013).
13. Woo, S. R. et al. Immune inhibitory molecules LAG-3 and PD-1 synergistically regulate T-cell function to promote tumoral immune escape. Cancer Res. **72,** 917–927 (2012).
14. Landsberg, J. et al. Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. Nature **490,** 412–416 (2012).
15. Matsushita, H. et al. Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature **482,** 400–404 (2012).
16. Cheung, A. F., Dupage, M. J., Dong, H. K., Chen, J. & Jacks, T. Regulated expression of a tumor-associated antigen reveals multiple levels of T-cell tolerance in a mouse model of lung cancer. Cancer Res. **68,** 9459–9468 (2008).
17. Fuertes, M. B. et al. Host type I IFN signals are required for antitumor CD8+ T cell responses through CD8α+ dendritic cells. J. Exp. Med. **208,** 2005–2016 (2011).
18. Hildner, K. et al. Batf3 deficiency reveals a critical role for CD8α+ dendritic cells in cytotoxic T cell immunity. Science **322,** 1097–1100 (2008).
19. Bedoui, S. et al. Cross-presentation of viral and self antigens by skin-derived CD103+ dendritic cells. Nature Immunol. **10,** 488–495 (2009).
20. Mollah, S. A. et al. Flt3L dependence helps define an uncharacterized subset of murine cutaneous dendritic cells. J. Invest. Dermatol. **134,** 1265–1275 (2014).
21. Aliberti, J. et al. CCR5 provides a signal for microbial induced production of IL-12 by CD8α+ dendritic cells. Nature Immunol. **1,** 83–87 (2000).
22. Peng, W. et al. PD-1 blockade enhances T-cell migration to tumors by elevating IFN-γ inducible chemokines. Cancer Res. **72,** 5209–5218 (2012).
23. Li, Y. et al. N-myc downstream-regulated gene 2, a novel estrogen-targeted gene, is involved in the regulation of Na+/K+-ATPase. J. Biol. Chem. **286,** 32289–32299 (2011).
24. Khuu, C. H., Barrozo, R. M., Hai, T. & Weinstein, S. L. Activating transcription factor 3 (ATF3) represses the expression of CCL4 in murine macrophages. Mol. Immunol. **44,** 1598–1605 (2007).
25. Jongbloed, S. L. et al. Human CD141+ (BDCA-3)+ dendritic cells (DCs) represent a unique myeloid DC subset that cross-presents necrotic cell antigens. J. Exp. Med. **207,** 1247–1260 (2010).
26. Spranger, S. et al. Mechanism of tumor rejection with doublets of CTLA-4, PD-1/PD-L1, or IDO blockade involves restored IL-2 production and proliferation of CD8+ T cells directly within the tumor microenvironment. J. Immunother. Cancer (2014).
27. Yaguchi, T. et al. Immune suppression and resistance mediated by constitutive activation of Wnt/β-catenin signaling in human melanoma cells. J. Immunol. **189,** 2110–2117 (2012).
28. Driessens, G. et al. β-Catenin inhibits T cell activation by selective interference with linker for activation of T cells–phospholipase C-γ1 phosphorylation. J. Immunol. **186,** 784–790 (2011).
29. Cipponi, A., Wieers, G., van Baren, N. & Coulie, P. G. Tumor-infiltrating lymphocytes: apparently good for melanoma patients. But why? Cancer Immunol. Immunother. **60,** 1153–1160 (2011).

METHODS

Analysis of TCGA data set. Level 4 gene expression data and level 2 somatic mutation data were downloaded for skin cutaneous melanoma (SKCM) from TCGA, which were processed by Broad Institute's TCGA workgroup (release date 10 October 2013). The RNA-seq level 4 gene expression data contain upper-quartile-normalized and $\log_2$-transformed RNA-seq by expectation maximization (RSEM) values summarized at gene level[30]. The whole-exome sequencing (WXS) level 2 mutation data contains somatic mutation calls for each subject. A total of 266 metastatic SKCM samples were analysed. For clustering of cold and hot tumours, genes expressed in less than 80% of the samples were removed. A total of 15,974 genes were kept for further analysis. Unsupervised hierarchical clustering of the genes was performed in primary tumours and metastasis samples separately using K-mean equal to 12 and Euclidean distance metrics. Clusters containing the 13 known T-cell-signature transcripts (*CD8A*, *CCL2*, *CCL3*, *CCL4*, *CXCL9*, *CXCL10*, *ICOS*, *GZMK*, *IRF1*, *HLA-DMA*, *HLA-DMB*, *HLA-DOA*, *HLA-DOB*) were selected for resampling-based hierarchical clustering of the samples using ConsensusClusterPlus v.1.16.0 (ref. 31). This procedure was performed with 2,000 random selections of 80% of the samples and Euclidean distance metrics. Genes differentially expressed between cold and hot tumour groups were detected using ANOVA and filtered by false discovery rate (FDR) q value < 0.01 and fold change > 2.0. Canonical pathways significantly enriched in the genes of interest were identified by Ingenuity Pathways Analysis (IPA) (Ingenuity Systems; http://www.ingenuity.com) based on experimental evidence from the Ingenuity Knowledge Base (release date 23 March 2014). The somatic variants were converted to VCF format and annotated using ANNOVAR (release date 23 August 2013)[32]. Each variant was annotated with known genes, exonic functions, predicted amino acid changes and minor allele frequencies derived from the 1000 Genomes Project (phase 1, release v.3, 23 November 2010) and the NHLBI Exome Sequencing Project (ESP6500SI-V2-SSA137) (EVS)[33]. Synonymous single-nucleotide variants (SNVs) were excluded from further analysis. The variants were then summarized at gene level and patient level for comparison of mutation profiles between the cold and hot tumour groups. Interactions between proteins encoded by genes of interest were retrieved from the STRING database based on high-confidence evidence collected from co-expression data, experiments and databases[34]. SNVs located in selected genes were analysed using the Variant Effect Prediction (http://www.ensembl.org/info/docs/tools/vep/index.html) software in combination with the UniProt database (http://www.uniprot.org). Calls of loss-of-function and gain-of-function were based on existing experimental data obtained from the UniProt data base, while harmful or tolerated effects on the protein structure were predicted using the SIFT prediction algorithm imbedded in the Variant Effect Predictions analysis. A continuous numerical score was generated using the six β-catenin target genes (*EFNB3*, *APC2*, *TCF1*, *c-MYC*, *TCF12*, *VEGFA*) reads. The resulting score was used to align patients based on activity of the β-catenin pathway.

Mice, tumour induction and generation of tumour cell lines. The following mouse strains were gifts from collaborators and were used to generate the mouse models used in this study: *Tyr:Cre-ER* (gifted by L. Chin), *LSL-Braf*^{V600E} (provided by M. MacMahon), *Pten*$^{fl/fl}$ (provided by T. Mak), *LSL-CAT-STA* (provided by F. Gounari), *Rosa26*-LSL-SIY and *Rosa26*-LSL-YFP (Jackson Laboratories, strain 006148) reporter[16,35–38]. As an initial cross, the *Tyr:Cre-ER* mice were crossed onto *LSL-Braf*^{V600E} and subsequently crossed with *loxP-Pten* mouse strain. Those mice were maintained as *Tyr:Cre-ER*$^+$, *LSL-Braf*$^{V600E+/−}$, *Pten*$^{fl/fl}$ and will be referred to as *Braf*^{V600E}/*Pten*$^{−/−}$. Additionally *Tyr:Cre-ER*, *LSL-Braf*^{V600E} mice were crossed to the *LSL-CAT-STA* mouse strain with subsequent crossing to the *Pten*$^{fl/fl}$ strain. Those mouse strains were maintained as *Tyr:Cre-ER*$^+$, *LSL-Braf*$^{V600E+/−}$, *LSL-CAT-STA*$^{+/+}$ and *Tyr:Cre-ER*$^+$, *LSL-Braf*$^{V600E+/−}$, *Pten*$^{fl/fl}$, *LSL-CAT-STA*$^{+/+}$ and will be referred to as *Braf*^{V600E}/*CAT-STA* or *Braf*^{V600E}/*Pten*$^{−/−}$/*CAT-STA*, respectively. Additionally, the *Braf*^{V600E}/*Pten*$^{−/−}$ and *Braf*^{V600E}/*Pten*$^{−/−}$/*CAT-STA* mice were crossed to the *Rosa26-LSL*-SIY mouse and mice were maintained heterozygote for the *Rosa26* locus. Similarly, *Braf*^{V600E}/*Pten*$^{−/−}$ mice were bred onto the *Rosa26-LSL*-YFP reporter strain, which were also maintained as heterozygous breeders for this locus. Genotyping was performed as described previously[16,35–39] (for primer sequences, see Supplementary Table 5). For tumour induction, 6–10-week-old mice were shaved on the back and 5 μl of 4-OH-tamoxifen (Sigma) at a concentration of 10 mg ml$^{−1}$ (dissolved in acetone) were applied. Subsequently, mice were screened weekly for tumour induction and growth with endpoint criteria of 4,000 mm^3. For tumour cell line generation, a single-cell suspension of the tumour tissue was generated as described later and as its entirety used for subcutaneous injections into Rag-knockout mice (RAGN12-F; Taconic). After tumour outgrowth, the tumour tissue was harvested and re-injected into *Rag*-knockout mice, C57BL/6 mice (Taconic), and adapted to cell culture using DMEM (Gibco) with 10% FCS (Atlanta Biologics), 1× NEAA (Gibco) and 1× MOPS (Sigma). In this work we used one cell line derived from each genotype, *Braf*^{V600E}/*Pten*$^{−/−}$ and *Braf*^{V600E}/*Pten*$^{−/−}$/*CAT-STA*. Additionally, TCR-transgenic 2C T cells were maintained as T-cell donors[40], actin–GFP mice were obtain from Jackson (strain identifier 003291), *Batf3*$^{−/−}$ mice were maintained as bone marrow donors and were originally obtained from K. Murphy[18]. All animal procedures were approved by the Institutional Animal Care and Use Committee of the University of Chicago. Human tumour cell lines were obtained from National Cancer Institute and maintained in RPMI medium supplemented with 10% FCS and 1× NEAA.

Tumour growth, tissue harvest and single-cell suspensions. For tumour outgrowth experiments, mice were treated at the lower back with 4-OH-tamoxifen at day 0. After day 21, tumour masses were measured by assessing length, width and height of major tumour mass using a digital calliper. Measuring the height was a critical parameter to assess tumour growth, since width and length were mainly influenced given by the spread of the TAM solution. Tumour volume T_V was calculated: $T_V = T_L \times T_W \times T_H$, where T_L is tumour length, T_H is tumour height and T_W is tumour width, since the tumour shape was rectangular and flat rather than spherical. The maximum tumour size was reached when the tumour mass reached approximately 10% of the body weight. At the indicated experimental endpoint, tumour tissue was harvested, cleared from remaining skin and minced using razor blades. Subsequently, tumour pieces were digested using the human tumour digestion kit (Miltenyi) in combination with the tissue dissociater (Miltenyi). For flow cytometric analysis and cell sorting, living cells were separated using a ficoll (GE) centrifugation step with subsequent washing of the obtained cells. For generation of tumour cell lines, the cell suspension was used directly after digestion and two washing steps.

Immunohistochemistry and fluorescent immunohistology. The immunohistology staining on human samples was performed by the Human Tissue Resource Center of the University of Chicago using biopsies from malignant melanoma patients. Staining was performed using a CD8-specific monoclonal antibody (CD8 clone C8/144B, NeoMarkers), β-catenin (clone CAT-5H1, Life Technologies) in combination with a secondary goat anti-mouse immunoglobulin G (IgG) conjugated to an alkaline phosphatase (Biocare Medical) was applied. Slides were scanned using a CRi Panoramic Scan Whole Slide Scanner. Positivity for β-catenin staining was obtained first and grading was based on the staining intensity. Subsequently, the number of CD8-positive T cells within one needle biopsy (2.5 mm diameter) was counted using ImageJ cell counter and calculated as number of CD8$^+$ T cells per mm^2. Samples with fewer than 50 CD8$^+$ T cells per mm^2 were considered T-cell-infiltrate low whereas counts >50 per mm^2 were considered as T-cell high, similar to as described previously[41]. For mouse fluorescent immunohistology staining, formalin/paraffin-fixed tissues were used to obtain 5 μm sections for subsequent staining. Staining was performed using the following primary antibodies: anti-CD3 (clone SP7, 1:500, Abcam) and anti-Trp1 (clone EPR13063, 1:500, Abcam) in combination with goat anti-rabbit 594 (JacksonImmuno) and Hoechst counterstain. Slides were imaged using a Zeiss Axiovert 200 with a Hammatsu Orca ER firewire digital monochrome camera.

Flow cytometry and cell sorting. For flow cytometric analysis, washed cells were resuspended in staining buffer (PBS with 10% FCS and 0.5 M EDTA (Ambion)). Cells were incubated with live/dead staining dye (Invitrogen, wavelength 450 nm) and Fc Block (clone 93; Biolegend) for 20 min on ice. Subsequently, specific antibodies were added (Supplementary Table 5) and staining was continued for 40 min on ice. After a washing step, cells were either analysed directly or fixed with 4% PFA (BD) solution for 30 min and stored in a 1% PFA solution until analysis. For staining of TCR-transgenic 2C T cells a TCR specific-biotinylated monoclonal antibody (1B2 clone) was obtained from the University of Chicago Monoclonal Core Facility. Subsequent to live/dead staining, TCR-specific monoclonal antibody was added for 15 min on ice at a 1:100 dilution alone with surface antibodies targeting other antigens added in for an additional 25 min thereafter. After a washing step, a 1:500 dilution of Streptavidin APC was added and incubated on ice for 20 min before cells were fixed in 4% PFA and stored in 1% PFA solution. Flow cytometry sample acquisition was performed on a LSR2B (BD), and analysis was performed using FlowJo software (TreeStar). For cell sorting, staining protocols were carried out similarly under sterile conditions. Cell sorting was performed using an ARIAIIIu (BD) and cells were collected in 100% FCS if further used for *in vitro* analysis or in TriZol Reagent (Invitrogen) if used for RNA isolation. Percentage of T cells was calculated as follows ((100/number of total living cells acquired) × number of CD3$^+$ T cells); number per gram tumour was calculated as follows (number of acquired CD3$^+$ T cells/tumour weight).

T-cell stimulation. 2.5×10^4 sorted T cells from spleen and/or tumour were either stimulated on plates coated with $1\,\mu g\,ml^{-1}$ anti-CD3 antibody (145-2C11 clone; Biolegend) and $2\,mg\,ml^{-1}$ anti-CD28 antibody (37.51 clone BD) in T-cell medium (DMEM, 10% FCS, $1\times$ NEAA, $1\times$ MOPS, $500\,\mu M$ β-mercapthoethanol (Sigma)) or plated on tissue-culture-treated uncoated plates for 8 h. Following incubation, cells were harvested and resuspended in TriZol Reagent (Invitrogen) for subsequent RNA isolation.

RNA isolation and qRT–PCR. RNA isolation using TriZol was performed according to the manufacturer's instruction. In the case of RNA isolation from whole tumour tissue, a piece of tumour was snap frozen in TriZol at the time of tumour harvest. Before RNA isolation the tissue was thawed at room temperature and homogenization was achieved using a tissue homogenizer (GE) with homogenizer tips (USA Scientific). Subsequent RNA isolation was performed according to the manufacturer's instructions. Reverse transcriptase reaction was performed using High Capacity cDNA RTPCR Kit (Life Technologies) according to instructions and 1 μl of the resulting copy DNA was used for qPCR. qPCR reactions were carried out using Sybr Green or TaqMan master mix (Life Technologies) and defined primer sets or primer/probe sets (probes were obtained from Roche), respectively (Supplementary Table 5). Reactions were run on a 7300 RT PCR system machine (Applied Biosystems) and expression level and fold change were calculated as follows: $\Delta CT = CT_{gene\ of\ interest} - CT_{18S}$; expression level $= 2^{-\Delta CT}$; fold change $= 2^{(\Delta CT_{reference\ sample} - \Delta CT_{tested\ sample})}$ (ref. 42).

Adoptive T-cell transfer. For adoptive transfer experiments, tumour development was induced and transfer of 1×10^6 T cells was performed when tumour reached near endpoint sizes (approximately 3–4 weeks after induction). Transferred T cells were isolated from gender-matched 2C donor mice using the Miltenyi CD8$^+$ enrichment Kit II for untouched CD8$^+$ T-cell isolation. After isolation, cells were stained with $1\,\mu M$ CFSE solution (eBioscience) for 8 min at 37 °C before intravenous injection. Tumour tissue, tumour-draining lymph nodes and spleen were harvested 5 days after adoptive transfer of T cells and used for flow cytometric analysis. This short timeframe was chosen to avoid the reported leakiness of the SIY transgene that has been associated with partial T-cell activation within the spleen[16]. For tumour tissues, the entirety of each sample was acquired and the total number of CD3$^+$CD8$^+$ T cells and transferred 2C cells was assessed. The percentage 2C cells was calculated as $((100/CD3^+/CD8^+\ T\ cells) \times 2C)$ and also the number of 2C cells per gram tumour.

Generation of bone marrow chimaeras. To condition host mice to generate bone marrow chimaeras, indicated mouse strains were irradiated twice with a 3 h interval and a first irradiation dose of 500 rad followed by 550 rad. Twenty-four hours after the second irradiation dose, bone marrow from gender-matched donor mice was isolated from femur and tibia of both legs, washed, and erythrocytes were lysed. 3×10^6 bone marrow cells were injected intravenously to reconstitute the mice. Two-to-three months after bone marrow transfer, tumour development was induced as described previously.

Generation and administration of bone-marrow-derived dendritic cells. For administration of bone-marrow-derived dendritic cells, bone marrow from C57BL/6 mice or GFP–actin mice was collected from the femurs and tibias of both legs. After washing and lysis of erythrocytes, bone marrow cells were cultured in RPMI (Gibco) complete medium (10% FCS, $1\times$ NEAA, $500\,\mu M$ β-ME) supplemented with $300\,ng\,ml^{-1}$ Flt3 ligand (eBioscience) for 7 days at a concentration of 2.5×10^6 cells ml^{-1}. Dendritic cells were then activated for 24 h with poly(I:C) (InvivoGen) at a final concentration of $5\,\mu g\,ml^{-1}$ (pre-heated for 5 min at 95 °C). Activated Flt3 ligand dendritic cells were frozen in aliquots of 5×10^6 cells in 90% FCS with 10% dimethylsulfoxide (DMSO; Sigma) until use for in vivo administration. For each dendritic cell preparation, activation marker expression was analysed using flow cytometry with the majority of cells being CD11c$^+$, CD11b$^+$, predominantly CD8α$^+$ and after activation high expression of CD80, CD86, MHCII and CD40 was observed. Injection of dendritic cells was initiated when the first signs of tumour lesions were identified on mice (2–3 weeks after induction) and were given intra-dermally/intra-tumourally using a 27G (Braintree) needle twice per week at a dose of 1×10^6 dendritic cells per injection.

Gene array analysis of mouse tumour tissue. For gene array analysis, RNA from whole tumour tissue was isolated. Subsequent experimental procedures were performed by the University of Chicago Genomics Core facility using the Illumina MouseWG-6 gene array chip (Illumina) according to the manufacturer's instructions. Subsequent gene lists were analysed from differentially expressed genes with a cut off for at least twofold change between the two analysed cohorts. Significance was determined using a two-way ANOVA test.

Trans-well migration assay. Dendritic cell populations were isolated from lymph nodes and skin of naive 6-week-old C57BL/6 mice. For this purpose, skin tissue was digested in a similar way as tumour tissue and cells from skin and lymph node were stained using the previously described protocol for cell sorting. Subsequently, living, CD45$^+$, CD11c$^+$, CD8α$^-$ or CD8α$^+$ cells were isolated from the lymph node sample as well as living, CD45$^+$, CD11c$^+$ or CD103$^+$ cells from skin samples. Migration assays were performed as described previously with minor adaptations using 5×10^5 cells per well and pre-treatment of dendritic cells with pertussis toxin (Sigma) at a final concentration of $20\,ng\,ml^{-1}$ for 1.5 h as indicated[43]. As a migration stimulus, CCL4 (R&D) was added to RPMI complete medium at 500 ng ml^{-1} or 48 h conditioned media from *Braf^{V600E}/Pten$^{-/-}$* or *Braf^{V600E}/Pten$^{-/-}$/CAT-STA* cell lines were used. At the endpoint, cells from the lower compartment as well as the trans-well were harvested and counted using a standard Neubauer counting chamber. Percentage of migrated cells was calculated as follows: count lower well/(count upper well + count lower well) $\times$ 100; with the sum of trans-well and lower well being >90% of the input cell count.

ELISA. ELISA assays against murine and human CCL4 were performed using CCL4-specific ELISA kits (R&D) according to the manufacturer's instructions.

siRNA knockdown. Target gene-specific and control siRNAs were obtained from Ambion and can be found in Supplementary Table 5. For knockdown, 3×10^4 tumour cells are plated in 96-well plates at a concentration of 3×10^5 per ml. Opti-MEM (Gibco) was mixed with 1.2 pmol siRNA and 1.5% RNAiMAX reagent (Invitrogen) and added to the culture at a ratio of 1:5. Cells were incubated for 48 h before supernatant was harvested for ELISA assays and cells were collected for RNA or protein extraction.

Western blotting. Cell lysates were generated using RIPA buffer in combination with protein inhibitor (Invitrogen) and protein concentration was determined using Bradford protein assay (Biorad). Denatured lysates were applied to a 10% SDS–PAGE and blotted using standard procedures. For protein detection, blots were incubated with primary antibodies (β-catenin clone D10A8; β-actin clone 13E5; Cell Signaling) overnight and with secondary antibodies (donkey anti-rabbit-HRP; GE Healthcare) for 2 h. Chemiluminescence was used to visualize the protein bands (GE Healthcare).

ChIP assay. For ChIP assays, the two cell lines, BP and BPC, were grown to 80% confluence in a 10 cm Petri dish. Cells were fixed with 1% formaldehyde solution for 30 min at 37 °C. Subsequent steps were performed using the EpiTect ChIP kit (Qiagen) according to the manufacturer's instructions with some minor adaptations. In brief, the formaldehyde was removed and cells were washed before harvesting using RIPA buffer. Sonication was performed using a water bath sonicator (GE) with the following cycle of 30 s on/15 s off at maximum voltage for 15 min, and this cycle was repeated three times at 4 °C. Chromatin-containing supernatants were incubated with an ATF3-specific antibody (mouse: polyclonal rabbit IgG; human: clone 44C3a, mouse IgG; Abcam) or rabbit/mouse IgG1 isotype (Cell Signaling) for 3 h or overnight at a 1:50 dilution. Pulled-down DNA was used as template for qPCR using Sybr Green master mix and primers (Supplementary Table 5). Results were calculated as followed: $\Delta CT = \Delta CT_{IP} - (\Delta CT_{IP} - \log_2{}^{100})$, fold enrichment $= 2^{(\Delta CT_{iso} - \Delta CT_{IP})}$.

Monoclonal antibody therapy. Therapy using monoclonal antibodies was initiated either when the tumour was first palpable or 7 days after dendritic cell injection was initiated. Mice were assigned to groups in a randomized fashion based on their ear tag number. Antibodies (CTLA-4 clone 9H10, PD-L1 clone 10F.9G2; BioXcell) were administered every other day throughout the experiment at a dose of 100 μg per mouse per treatment and treatment was initiated 3 weeks after tamoxifen application[26].

Statistical analysis. All statistical analyses were performed using GraphPad Prism (GraphPad) with the exception of analyses of the TCGA data set. Unless otherwise noted, all data are shown as mean $\pm$ s.e.m. combined with a two-tailed Mann–Whitney U test. Significance was assumed with $P \leq 0.05$. For correlation studies, a Gaussian fit was performed to assure normal distribution. All experiments shown were repeated at least in two independent experiments.

30. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
31. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26,** 1572–1573 (2010).
32. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
33. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
34. Jensen, L. J. et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37,** D412–D416 (2009).
35. Suzuki, A. et al. High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. *Curr. Biol.* **8,** 1169–1178 (1998).

36. Gounari, F. *et al.* Stabilization of β-catenin induces lesions reminiscent of prostatic intraepithelial neoplasia, but terminal squamous transdifferentiation of other secretory epithelia. *Oncogene* **21,** 4099–4107 (2002).

37. Dankort, D. *et al.* A new mouse model to explore the initiation, progression, and therapy of *BRAF^{V600E}*-induced lung tumors. *Genes Dev.* **21,** 379–384 (2007).

38. Bosenberg, M. *et al.* Characterization of melanocyte-specific inducible Cre recombinase transgenic mice. *Genesis* **44,** 262–267 (2006).

39. Jeong, J., Mao, J., Tenzen, T., Kottmann, A. H. & McMahon, A. P. Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.* **18,** 937–951 (2004).

40. Manning, T. C. *et al.* Antigen recognition and allogeneic tumor rejection in CD8$^+$ TCR transgenic/RAG$^{-/-}$ mice. *J. Immunol.* **159,** 4665–4675 (1997).

41. Erdag, G. *et al.* Immunotype and immunohistologic characteristics of tumor-infiltrating immune cells are associated with clinical outcome in metastatic melanoma. *Cancer Res.* **72,** 1070–1080 (2012).

42. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3,** 1101–1108 (2008).

43. Spranger, S. *et al.* Generation of Th1-polarizing dendritic cells using the TLR7/8 agonist CL075. *J. Immunol.* **185,** 738–747 (2010).
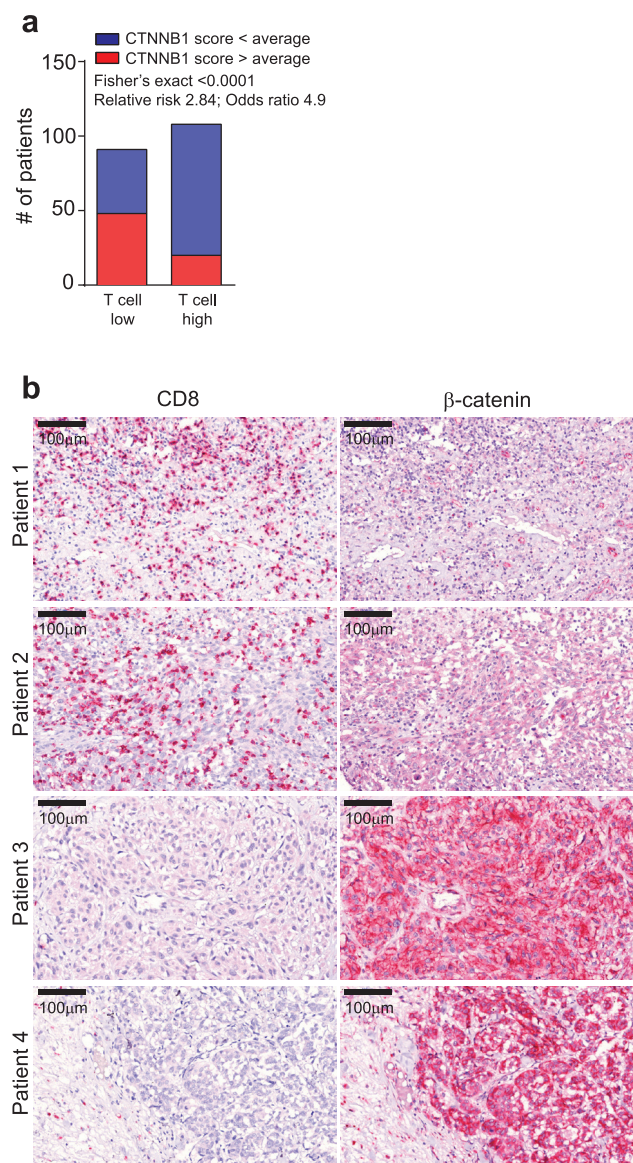
Extended Data Figure 1 | Correlation between active β-catenin and CD8 T-cell infiltrate in human patients. a, A continuous numerical score was generated using six β-catenin target genes (*CTNNB1* score). Using this score, patients from the TCGA data set were grouped in high or low *CTNNB1* score (centred on the average score) (low, 91 patients; high, 108 patients). Subsequent correlation analysis was performed using a Fisher's exact test. **b**, Representative examples for CD8 and β-catenin staining in human needle biopsies used for analysis shown in Fig. 1d.
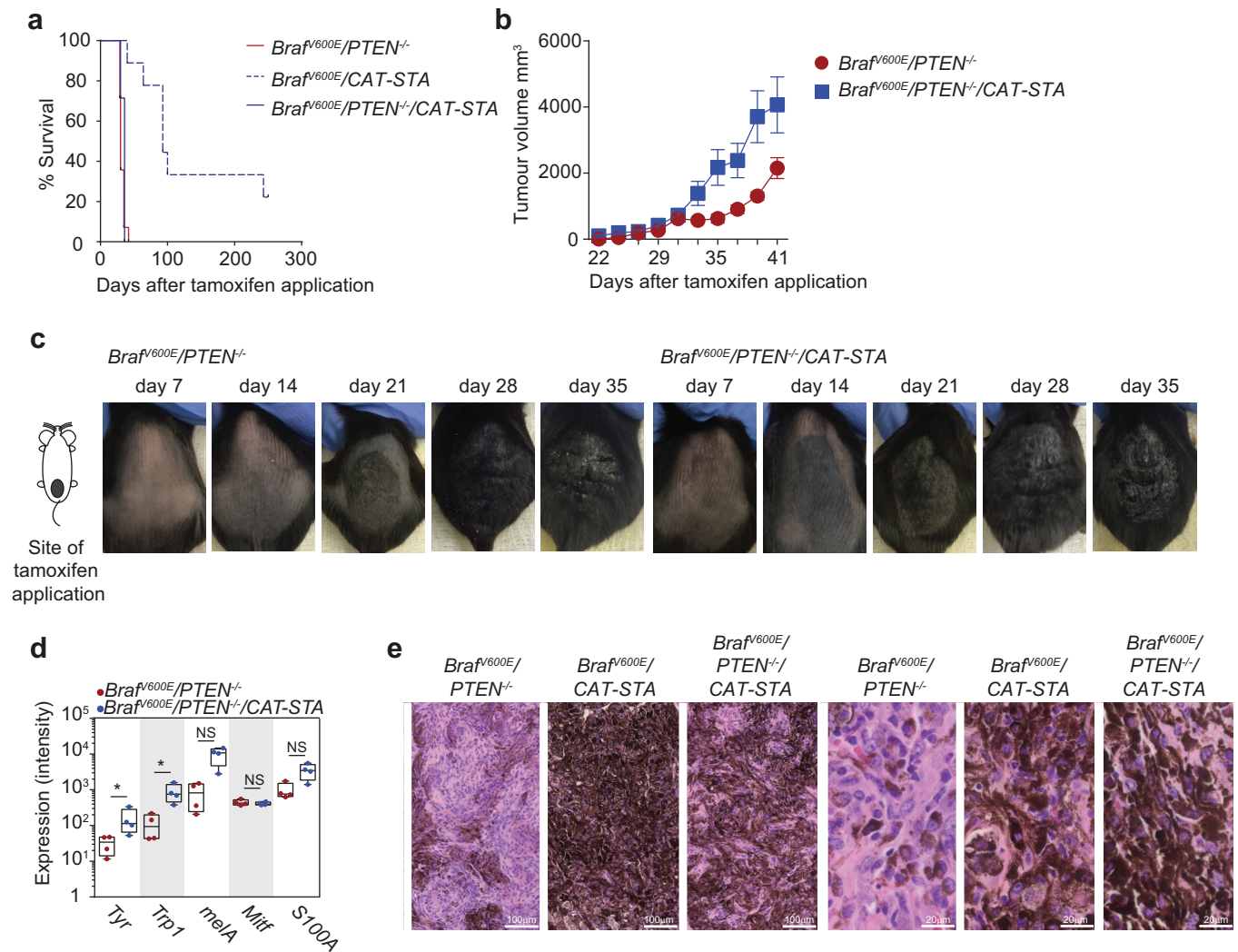
Extended Data Figure 2 | Tumour growth of genetically engineered mice.
a, Overall survival of all three models: $Braf^{V600E}/Pten^{-/-}$ with 100% lethality and mean time to death of 31 days ($n = 14$), $Braf^{V600E}/CAT$-STA with 85% lethality and mean time to tumour event of 93 days ($n = 8$), and $Braf^{V600E}/Pten^{-/-}/CAT$-$STA$ with 100% lethality and mean time to tumour event of 36 days ($n = 14$). **b**, Tumour outgrowth of $Braf^{V600E}/Pten^{-/-}$ (red) and $Braf^{V600E}/Pten^{-/-}/CAT$-$STA$ (blue) tumours shown as mm³ at days after tamoxifen application ($n = 10$). **c**, Representative macroscopic pictures for tumour growth over time when tamoxifen was applied on the lower back of the mouse (see illustration). **d**, Gene array analysis of tumours isolated from GEMs ($n = 4$, Mann–Whitney U test). **e**, Histology slides showing representative examples for haematoxylin and eosin stain in all three mouse models (left, ×20, scale bars indicate 100 μm; right, ×100, scale bars indicate 20 μm). *$P \leq 0.05$; NS, not significant.

a

b Braf^{V600E}/PTEN^{-/-}

c
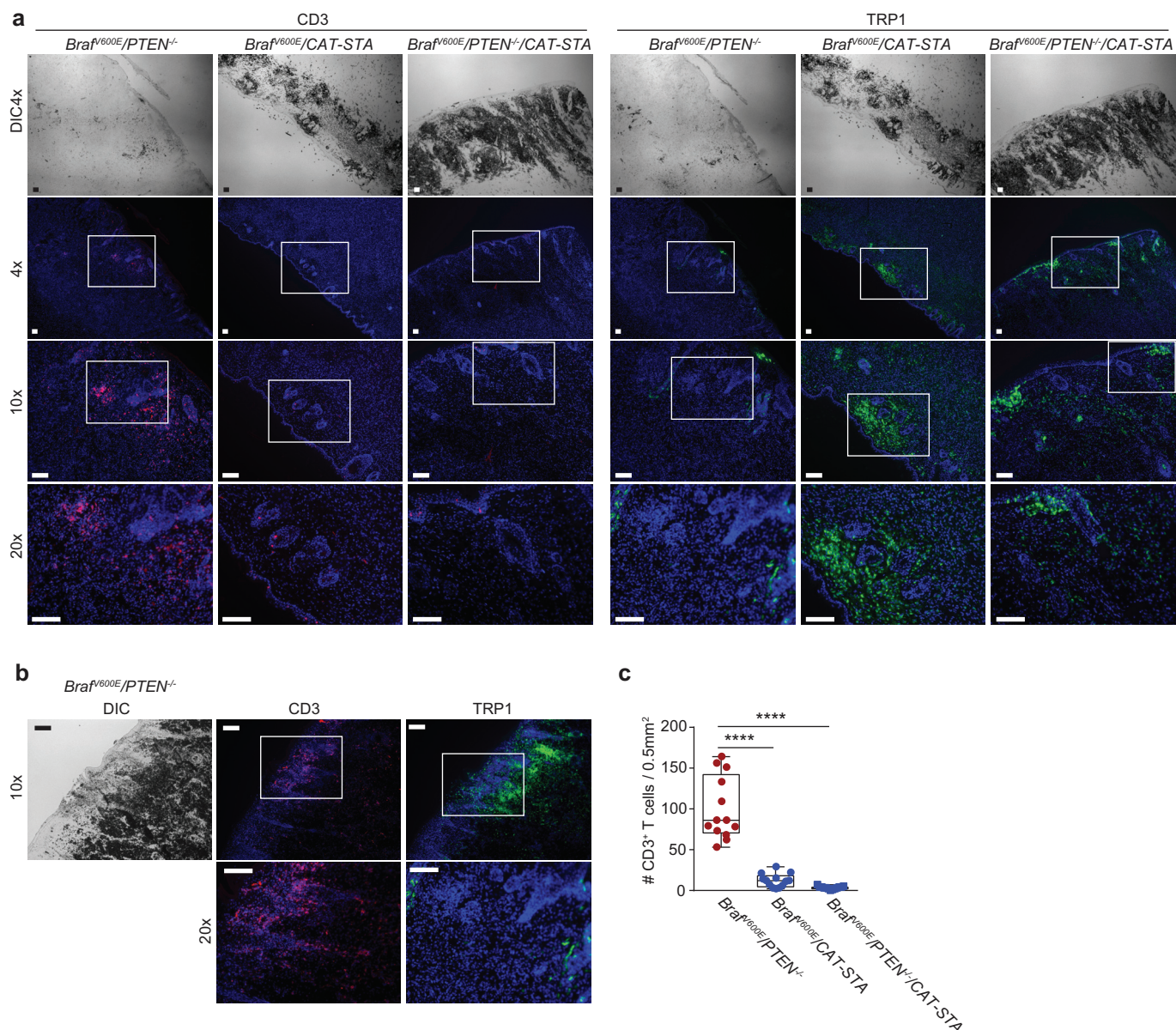
Extended Data Figure 3 | T-cell infiltration of genetically engineered mice. **a**, Representative images of immmunofluorescent staining against CD3 (red, left panel) and TRP1 (green, right panel) in all three tumour tissues (scale bar, 100 μm; ×4, ×10, ×20 with ×4 differential interference contrast (DIC) on top; nuclei Hoechst ×20 CD3 stain as shown in Fig. 1). **b**, Representative immmunofluorescent staining against CD3 (red, left panel) and TRP1 (green, right panel) in a highly pigmented area of Braf^{V600E}/Pten^{-/-} tumour tissues (scale bar, 100 μm; ×10, ×20 with ×10 DIC left) excluding that the lack of T cells is associated with increased pigmentation (nuclei Hoechst). **c**, Numbers of CD3^{+} T cells were counted within 13 different fields (0.5 mm × 1 mm) from two tumour samples. Mean of 12 T cells or 3.2 T cells per 0.5 mm^2 in Braf^{V600E}/CAT-STA or Braf^{V600E}/Pten^{-/-}/CAT-STA tumours, respectively, versus 100 T cells per 0.5 mm^2 in Braf^{V600E}/Pten^{-/-} tumours. Data are given as mean with minimum and maximum, as well as individual values. Statistical analysis was performed using Mann–Whitney U test. ****P ≤ 0.0001.
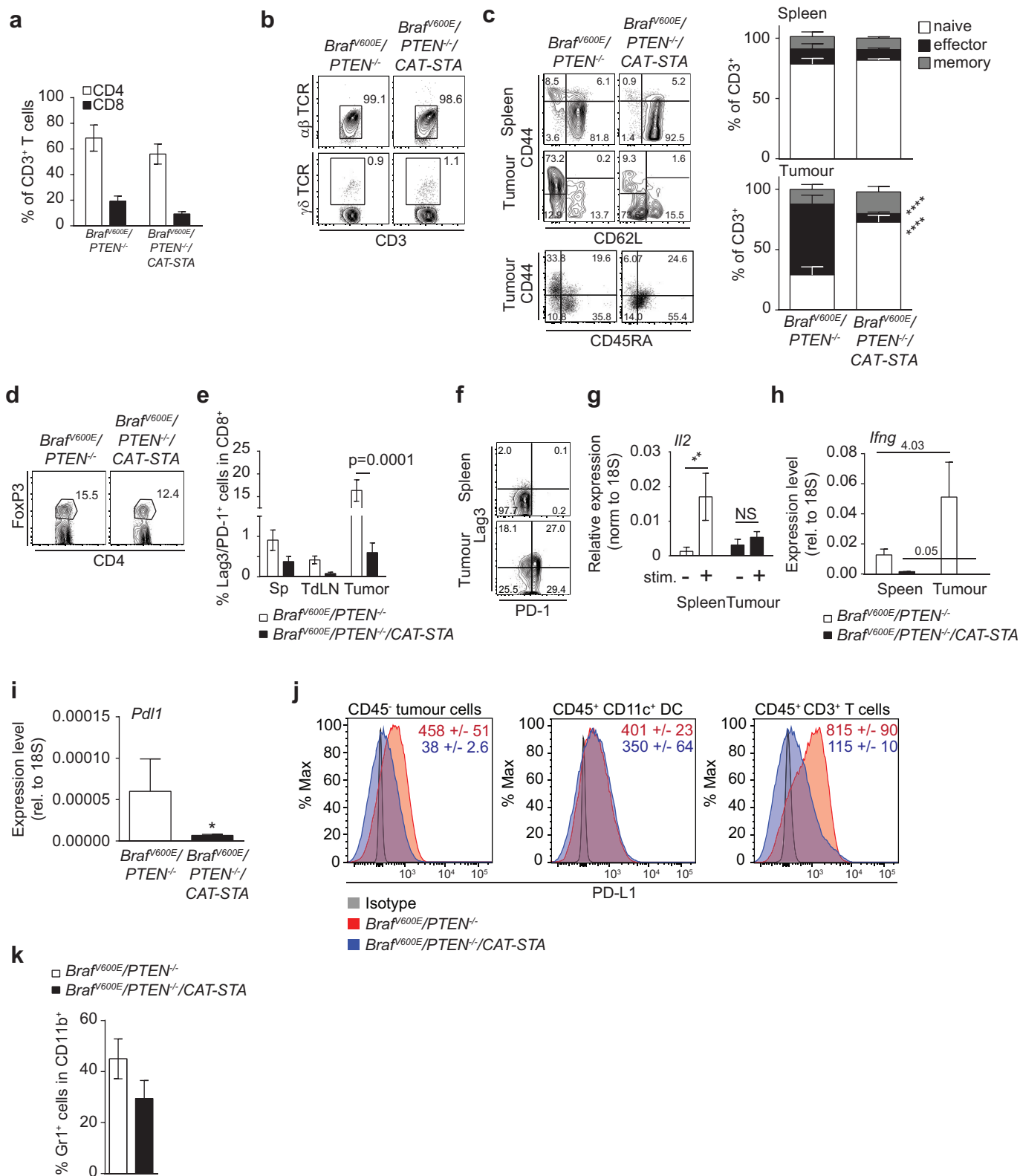
Extended Data Figure 4 | Characterization of the T-cell infiltrate in $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ mice. **a**, Distribution of T-cell subsets in $Braf^{V600E}/Pten^{-/-}$ and $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours ($n = 6$). **b**, **c**, Representative flow cytometry plots to discriminate αβ-TCR T cells and γδ-TCR T cells (**b**), naive (CD62L$^+$CD44$^-$) and effector (CD62L$^-$CD44$^+$) T cells (pre-gated on CD3$^+$CD8$^+$ T cells), and one representative example of CD44/CD45RA staining (**c**). Quantification of naive (CD62L$^+$CD44$^-$CD45RA$^+$), effector (CD62L$^-$CD44$^+$CD45RA$^-$) and memory (CD62L$^+$CD44$^+$CD45RA$^-$) T cells is indicated on the right ($n = 6$). **d**, Representative flow cytometry plots of FoxP3$^+$ T regulatory cells ($n = 6$). **e**, Quantification and comparison of PD-1/Lag3 double-positive T cells in $Braf^{V600E}/Pten^{-/-}$ and $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ tumours ($n = 12$). **f**, Representative flow cytometry of PD-1- and Lag3-positive T cells (pre-gated on CD3$^+$CD8$^+$ T cells) in $Braf^{V600E}/Pten^{-/-}$ tumours. **g**, $Il2$ transcripts present in sorted CD3$^+$ T cells from $Braf^{V600E}/Pten^{-/-}$ tumours and spleen ($n = 10$). **h**, $Ifng$ transcripts present in sorted CD3$^+$ T cells from $Braf^{V600E}/Pten^{-/-}$ and $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ mice ($n = 10$). **i**, Expression level of PD-L1 in whole tumour tissue from both mouse models assessed by qRT–PCR ($n = 8$). **j**, Flow cytometric analysis of PD-L1 expression of non-haematopoietic tumour cells (CD45$^-$), CD45$^+$CD11c$^+$ dendritic cells (DC) and CD45$^+$CD3$^+$ T cells. Shown is a representative example as histogram (grey isotype, red $Braf^{V600E}/Pten^{-/-}$; blue, $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$) with mean fluorescent intensity of $n = 3$ given each histogram (red, $Braf^{V600E}/Pten^{-/-}$; blue, $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$). **k**, Percentage of Gr1$^+$ cells within the CD11b$^+$ fraction of the tumour immune cell infiltrate ($n = 8$; absolute numbers $Braf^{V600E}/Pten^{-/-}$: $1,047 \pm 418$ cells per gram tumour to $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$: 739 ± 185 cells per gram tumour; $P = 0.7429$). All data are mean $\pm$ s.e.m., Mann–Whitney U test.*$P \le 0.05$, **$P \le 0.01$, ****$P \le 0.0001$; NS, not significant.
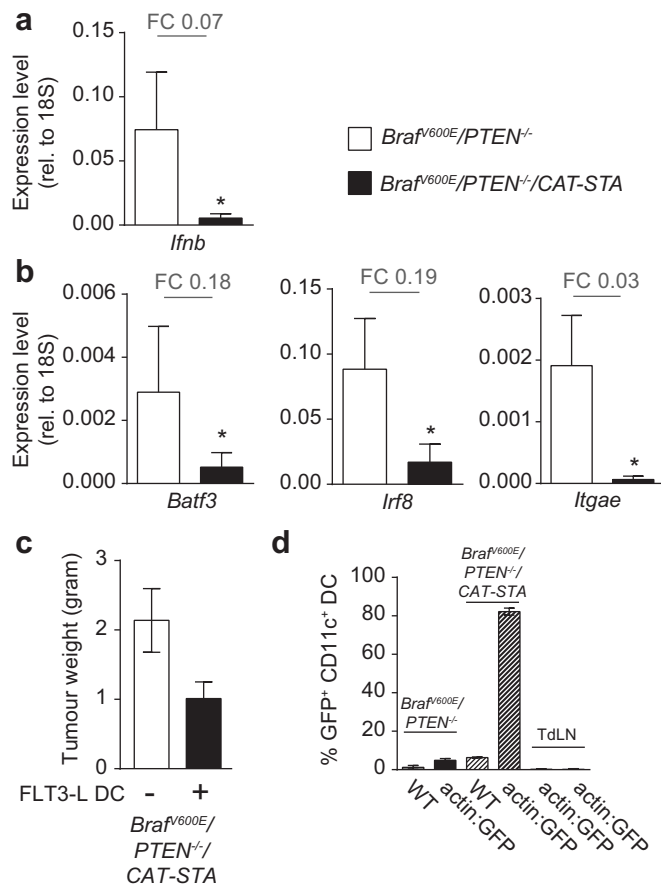
Extended Data Figure 5 | Injection of Flt3 ligand-derived dendritic cells into tumours of $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ mice is sufficient to overcome the lack of CD103$^+$ dermal dendritic cells. **a**, Expression level of *Ifnb* in CD45$^+$CD11c$^+$ sorted dendritic cells from tumours from $Braf^{V600E}/Pten^{-/-}$ (open bars) and $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ (filled bars) mice. FC, fold change. **b**, Expression level of *Batf3*, *Irf8* and *Itgae* in sorted dendritic cells. Fold change is indicated in each graph ($n = 8$). **c**, Mean ($\pm$ s.e.m.) tumour weight of $Braf^{V600E}/Pten^{-/-}/CAT\text{-}STA$ assessed at the endpoint of the experiment depicted in Fig. 3e, after intra-tumoural injection of dendritic cells. **d**, Per cent of GFP$^+$CD11c$^+$ dendritic cells (DC) present at the tumour site after injections of Flt3 ligand-derived dendritic cells from actin–GFP mice. Depicted are the percentages detected in the tumour of both genotypes injected with either wild-type or actin–GFP dendritic cells as well as in the TdLNs for the actin–GFP injected mice ($n = 4$). All data are mean $\pm$ s.e.m., Mann–Whitney *U* test. *$P \leq 0.05$.
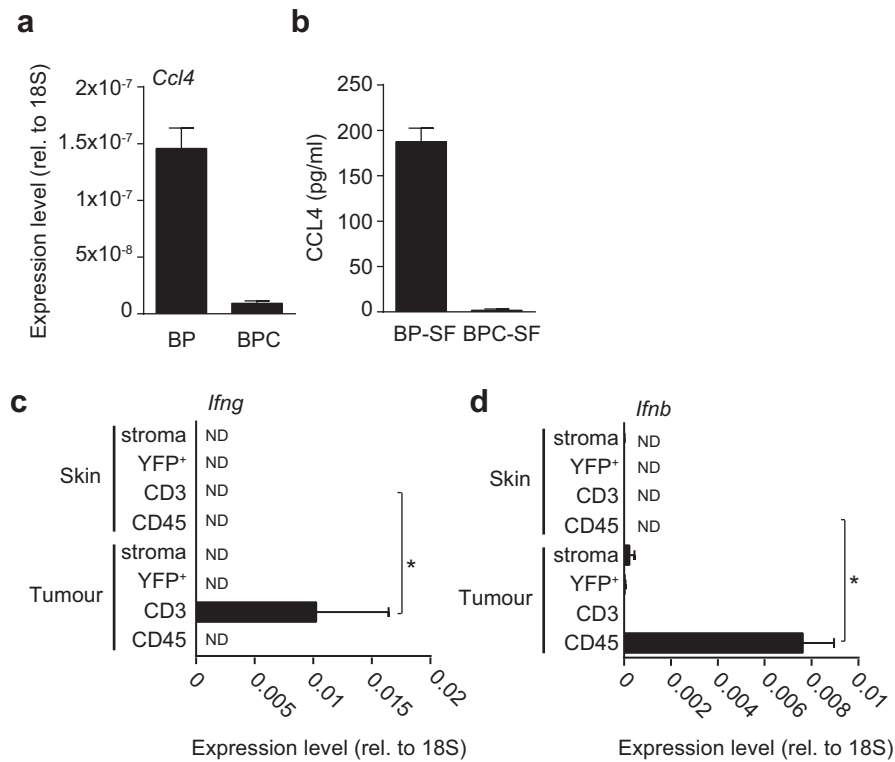
Extended Data Figure 6 | Chemokine expression patterns indicate that CCL4 expression from tumour cells is directly inhibited by active β-catenin-signalling. a, Expression of *Ccl4* mRNA in established tumour cell lines BP and BPC (8 independent experiments). **b**, Amount of secreted CCL4 in 48 h conditioned BP and BPC tumour cell supernatants, assessed by ELISA (4 independent experiments). **c, d**, Control qRT–PCR for the experiment shown in Fig. 4e with *Ifnb* expression (c) and *Ifng* expression (d) (*n* = 6). ND, not detected. All data are mean ± s.e.m., Mann–Whitney *U* test. *P ≤ 0.05.
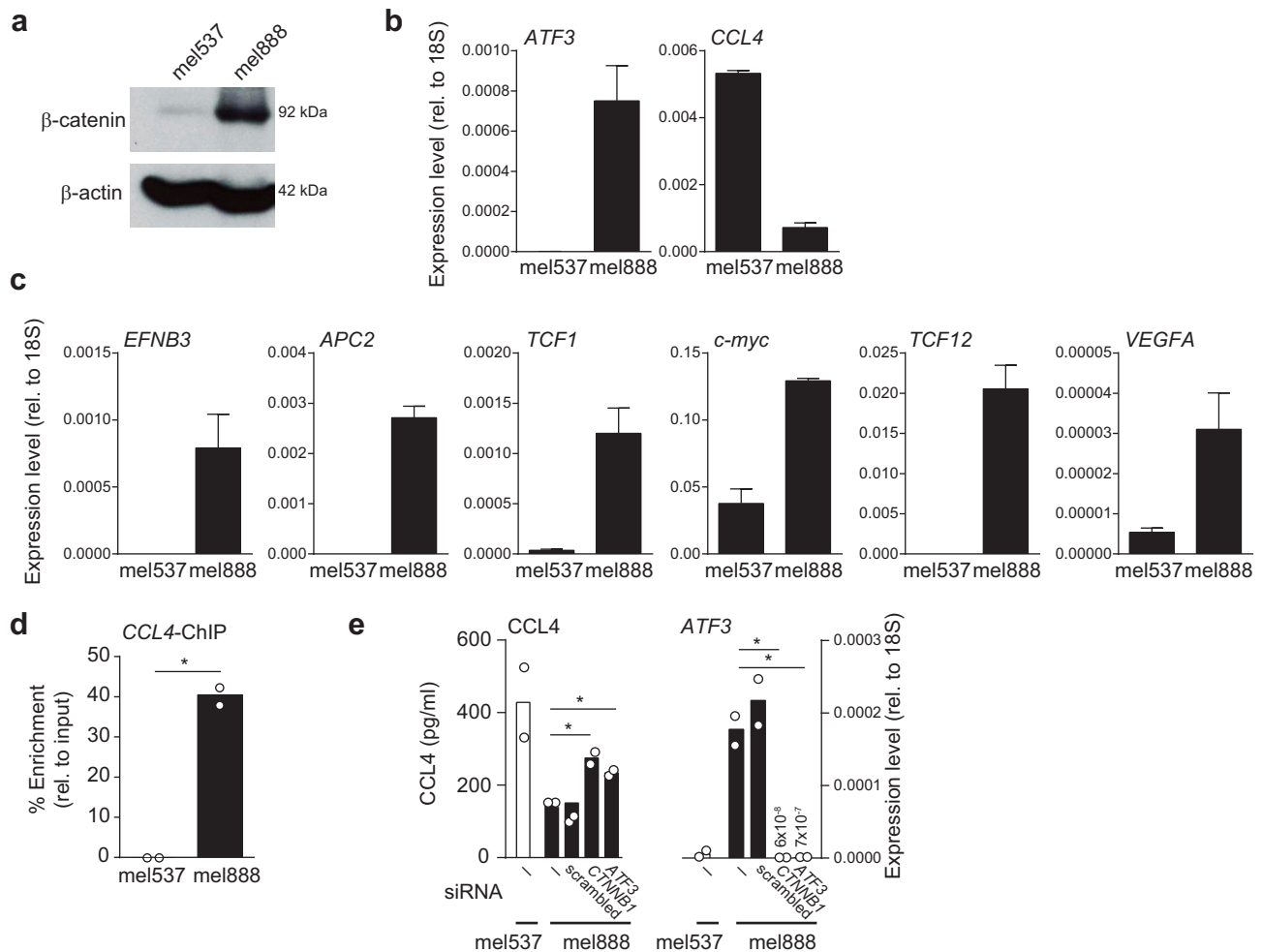
Extended Data Figure 7 | Active β-catenin signalling blocks CCL4 production in human melanoma cell lines. a, Western blot on mel537 and mel888 showing stabilized β-catenin expression. **b**, Expression level of human *ATF3* and human *CCL4* in mel537 and mel888 (three independent experiments, duplicates per experiment). **c**, Expression level of β-catenin target genes in mel537 and mel888. **d**, ATF3-specific ChIP assay in mel537 and mel888 cell lines for the *CCL4* gene locus (two independent experiments, duplicates per experiment). **e**, CCL4 secretion (left) and *ATF3* transcription levels (right) after siRNA-mediated knockdown of *CTNNB1 and ATF3* in mel537 and mel888 assessed by ELISA or qRT–PCR, respectively (two independent experiments, duplicates per experiment). All data are mean ± s.e.m., Mann–Whitney *U* test. *$P \leq 0.05$.
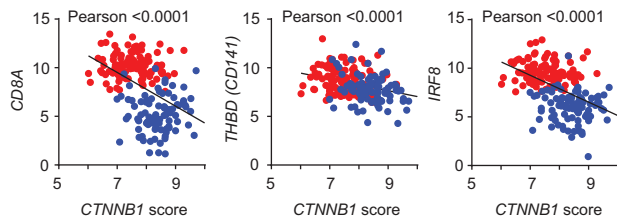
Extended Data Figure 8 | β-Catenin target gene expression correlates inversely with markers for human BATF3-lineage dendritic cells and T cells. Pearson correlation of *CTNNB1* score with *CD8A* ($R^2 = 0.214$), *THBD* ($R^2 = 0.109$) and *IRF8* ($R^2 = 0.2374$) (red indicates T-cell-signature high, blue indicates T-cell-signature low).
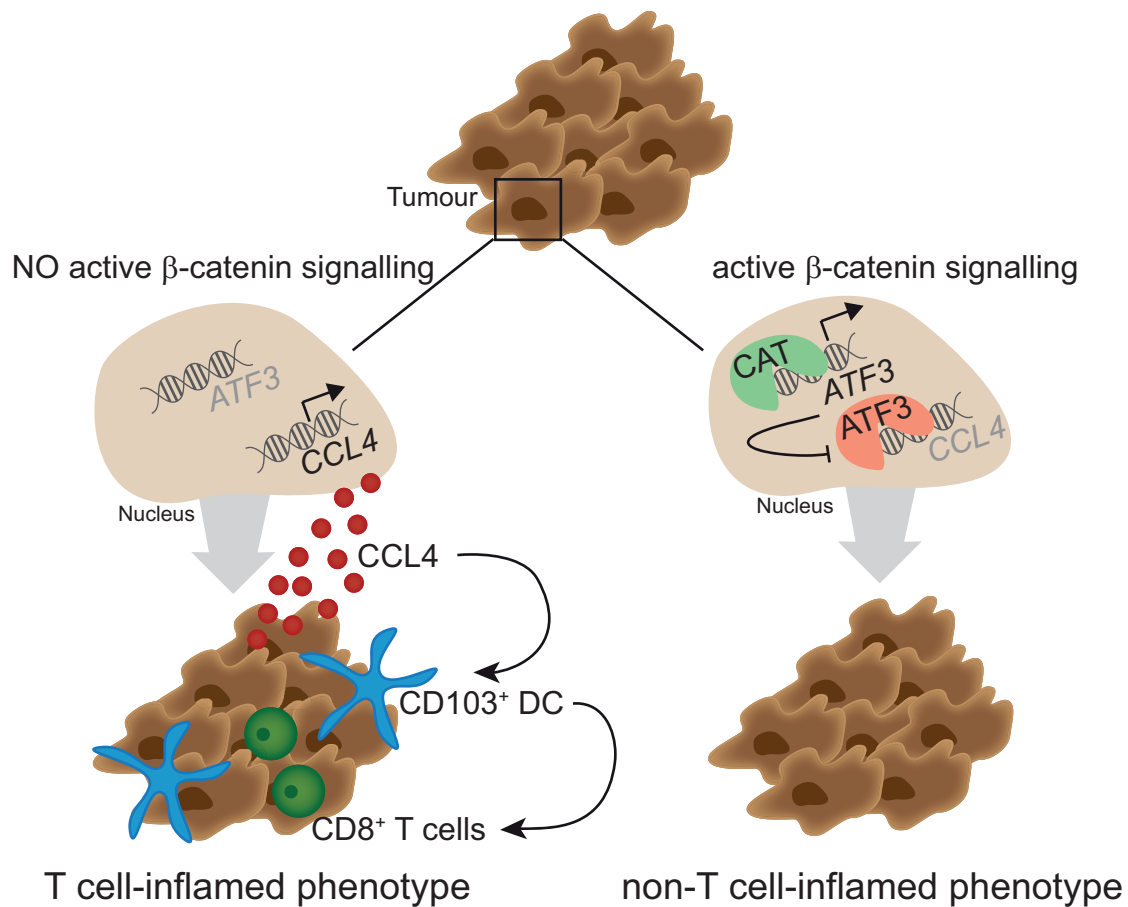
Extended Data Figure 9 | Graphical summary. Left, tumour without active β-catenin signalling in which *ATF3* transcription is not induced and thus *CCL4* (red circles) is transcribed and secreted. Downstream CD103[+] dendritic cells (DC) (blue) are attracted and subsequent activation of CD8[+] T cells (green) is enabled. Right, tumour with active β-catenin signalling (green), which leads to induction of *ATF3* transcription (red), which in turn leads, among others effects, to suppression of *CCL4* transcription. This leads to an active escape from the anti-tumour immune response since dendritic cell recruitment is insufficient.

LETTER

Cyclic di–GMP acts as a cell cycle oscillator to drive chromosome replication

C. Lori[1]*, S. Ozaki[1]*, S. Steiner[1]†, R. Böhm[2], S. Abel[1]†, B. N. Dubey[2], T. Schirmer[2], S. Hiller[2] & U. Jenal[1]

Fundamental to all living organisms is the capacity to coordinate cell division and cell differentiation to generate appropriate numbers of specialized cells. Whereas eukaryotes use cyclins and cyclin-dependent kinases to balance division with cell fate decisions[1], equivalent regulatory systems have not been described in bacteria. Moreover, the mechanisms used by bacteria to tune division in line with developmental programs are poorly understood. Here we show that *Caulobacter crescentus*, a bacterium with an asymmetric division cycle, uses oscillating levels of the second messenger cyclic diguanylate (c-di-GMP) to drive its cell cycle. We demonstrate that c-di-GMP directly binds to the essential cell cycle kinase CckA to inhibit kinase activity and stimulate phosphatase activity. An upshift of c-di-GMP during the G1–S transition switches CckA from the kinase to the phosphatase mode, thereby allowing replication initiation and cell cycle progression. Finally, we show that during division, c-di-GMP imposes spatial control on CckA to install the replication asymmetry of future daughter cells. These studies reveal c-di-GMP to be a cyclin-like molecule in bacteria that coordinates chromosome replication with cell morphogenesis in *Caulobacter*. The observation that c-di-GMP-mediated control is conserved in the plant pathogen *Agrobacterium tumefaciens* suggests a general mechanism through which this global regulator of bacterial virulence and persistence coordinates behaviour and cell proliferation.

To enable tissue homeostasis, metazoans tightly regulate the balance between cell proliferation and differentiation[2]. Cyclin-dependent kinases (CDKs) are particularly important in cell proliferation, development and cell fate decisions[1,3]. To drive cell cycle progression, CDKs associate with oscillating, stage-specific regulatory subunits called cyclins[4]. While in higher organisms cells generally undergo terminal differentiation, bacteria often rely on rapid growth to exploit available nutrients and thus need to dynamically tune behavioural programs with cell proliferation. How exactly bacteria couple behavioural processes with cell cycle progression remains unclear.

A prime model to study the coupling of cell growth and behaviour in bacteria is the aquatic organism *Caulobacter crescentus*, which strictly separates cell motility from cell proliferation. *C. crescentus* divides asymmetrically to generate two specialized progeny, a sessile and replication-competent stalked cell and a motile and replication-inert swarmer cell. The swarmer cell (G1 phase) re-enters the replication cycle during differentiation into a stalked cell (S phase) (Fig. 1a). To control the motile–sessile transition, *C. crescentus* makes use of c-di-GMP, a second messenger controlling a wide range of behavioural processes in bacteria, including virulence, motility and biofilm formation[5]. C-di-GMP levels are low in swarmer cells, increase during differentiation to peak in stalked cells and later reach intermediate levels in the pre-divisional cell[6]. One of the main drivers of c-di-GMP fluctuations is the diguanylate cyclase PleD, which is active in stalked but turned off in swarmer cells[7] (Fig. 1a). While a *pleD* mutant has reduced levels of c-di-GMP, a strain lacking all diguanylate cyclases (cdG0) is devoid of c-di-GMP[6]. The complete loss of

motility and surface attachment in the cdG0 strain illustrates the importance of c-di-GMP oscillation for *Caulobacter* cell fate determination[6]. In contrast, the role of c-di-GMP in cell cycle progression is unclear.

Our studies originated from a genetic screen for synthetic lethal mutants in the cdG0 background. This strain, although viable, shows pronounced morphology and cell cycle defects[6]. We thus reasoned that c-di-GMP controls cell cycle progression together with a parallel pathway with partial functional redundancy. The screen revealed a strain with a transposon (Tn) insertion in the promoter region of the gene encoding the single-domain response regulator DivK (P*divK*::Tn) (Extended Data Fig. 1a). Crossing back the transposon into the cdG0 mutant produced a strain with severe cell cycle defects (Fig. 1b and Extended Data Fig. 1b). This, and the observation that the DNA content per cell mass unit was severely reduced (Fig. 1c), indicated that cells are severely compromised for replication initiation. In contrast, growth, division and replication were not affected when the P*divK*::Tn was crossed into a cdG+ strain (Fig. 1b, c). DivK levels were reduced about tenfold in the cdG0 P*divK*::Tn strain (Extended Data Fig. 1c), suggesting that DivK may be limiting for growth. This was confirmed by replacing the *divK* promoter upstream of the *divK* gene with the xylose-dependent promoter P*xyl*. In the absence of the inducer or in the presence of glucose, which further represses P*xyl* activity, DivK levels were strongly reduced compared to wild type (Extended Data Fig. 1d), resulting in severely reduced growth and replication in the cdG0 strain, but not in a cdG+ background (Fig. 1b and Extended Data Fig. 1e). Together, this indicated that c-di-GMP and DivK convergently regulate cell cycle progression.

DivK was recently shown to downregulate the central cell cycle kinase CckA through a direct interaction with DivL, an unorthodox kinase that controls CckA through a protein–protein interaction[8,9]. CckA initiates a phosphorelay controlling the activity of the response regulator CtrA[10,11] (Fig. 1a). CtrA is phosphorylated and active in swarmer cells (G1) where it binds to the origin of replication to inhibit replication initiation[12]. During differentiation into stalked cells CtrA is inactivated to license replication initiation[13]. CckA is bifunctional and can act both as kinase and as phosphatase to control CtrA via the phosphotransfer protein ChpT[14]. Accordingly, switching CckA from kinase to phosphatase activity during G1–S would rapidly reverse the phosphate flux to inactivate CtrA and authorize replication initiation. Hence, we reasoned that DivK and c-di-GMP could cooperate to inactivate CtrA. Because c-di-GMP controls CtrA degradation during G1–S transition through the effector protein PopA[15] (Fig. 1a), the G1 arrest of the cdG0 P*divK*::Tn strain could conceivably result from simultaneous over-activation and stabilization of CtrA. However, a mutant combining the P*divK*::Tn allele with a *popA* deletion stabilizing CtrA was not affected in terms of growth or DNA replication. In contrast, a P*divK*::Tn Δ*popA* strain that also lacked PleD, produced a strong G1 arrest (Fig. 1c and Extended Data Fig. 1f). From this we concluded that c-di-GMP regulates both stability and phosphorylation levels of CtrA during the cell cycle (Fig. 1a).
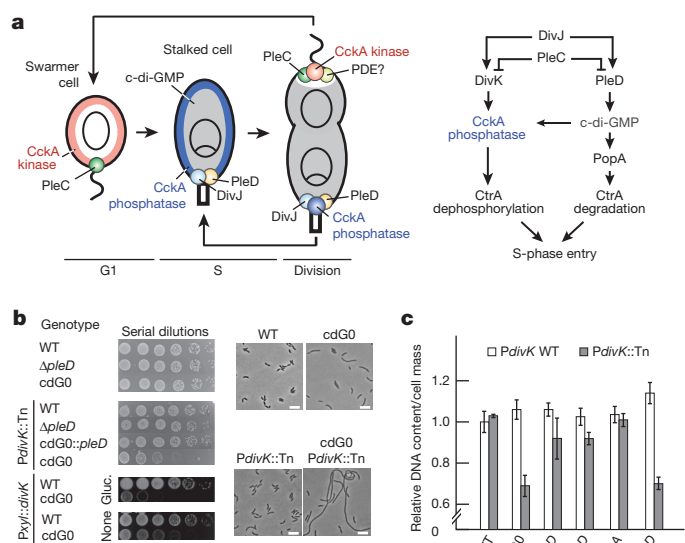
Figure 1 | C-di-GMP regulates cell cycle progression via the CckA-CtrA phosphorelay. **a**, Left, localization of CckA and factors regulating CckA activity throughout the *C. crescentus* cell cycle. CckA kinase (red) and phosphatase (blue) activities are indicated. High and low levels of c-di-GMP are shown as grey or white areas, respectively. PDE, phosphodiesterase. Right, Regulatory modules inactivating CtrA to control *Caulobacter* S-phase entry. **b**, Growth (left) and cell morphology (right) of strains indicated. Fivefold serial dilutions are shown. P*xyl*::*divK* strains were grown on peptone yeast extract (None) or peptone yeast extract supplemented with glucose (Gluc.) plates. Representatives of two biological replicates are shown. WT, wild type. **c**, Effect of P*divK*::Tn on DNA replication in different genetic backgrounds. DNA content and cell mass were determined by flow cytometry. Mean and s.d. for DNA content/cell mass were obtained from 4 biological replicates.

To analyse how c-di-GMP regulates CtrA activity, individual components of the CckA–CtrA phosphorelay were purified and examined *in vitro*. In the absence of c-di-GMP CckA autophosphorylation and phosphotransfer via ChpT to CtrA were readily observed. Notably, the addition of c-di-GMP completely abolished phosphorylation of all three components (Extended Data Fig. 2a). When CckA autophosphorylation was first carried out in the absence of c-di-GMP followed by the addition of c-di-GMP to the reaction mixture, rapid dephosphorylation of CckA was observed, suggesting that c-di-GMP is a potent stimulator of CckA phosphatase activity (Fig. 2a and Extended Data Fig. 2b). Stimulation of the CckA phosphatase was specific to c-di-GMP with GMP, GTP or cGMP having no observable effect (Fig. 2a). Experiments with all three components of the phosphorelay demonstrated that c-di-GMP effectively reverses the phosphate flux of the phosphorelay leading to the inactivation of CtrA (Fig. 2b). To test if c-di-GMP also regulates CckA kinase activity we compared phosphorylation *in vitro* of wild-type CckA with CckA(V366P), a mutant lacking phosphatase activity (Extended Data Fig. 2c)[14]. When c-di-GMP was added together with [^{32}P]ATP at the reaction start, CckA(V366P) phosphorylation was strongly reduced as compared to a control lacking c-di-GMP (Extended Data Fig. 2c), indicating that c-di-GMP inhibits CckA kinase activity.

These experiments demonstrate that c-di-GMP is a potent trigger to switch CckA from its default kinase into the phosphatase state. Consistent with this, purified CckA specifically binds radiolabelled c-di-GMP (Fig. 2c and Extended Data Fig. 2d). Further studies exposed the catalytic ATP-binding domain (CA) as minimal binding region for c-di-GMP (Extended Data Fig. 2e–g). To identify amino acid residues of the CA domain that are specifically involved in c-di-GMP binding, we concentrated on a candidate mutation that was recently isolated in the CckA homologue of the plant pathogen
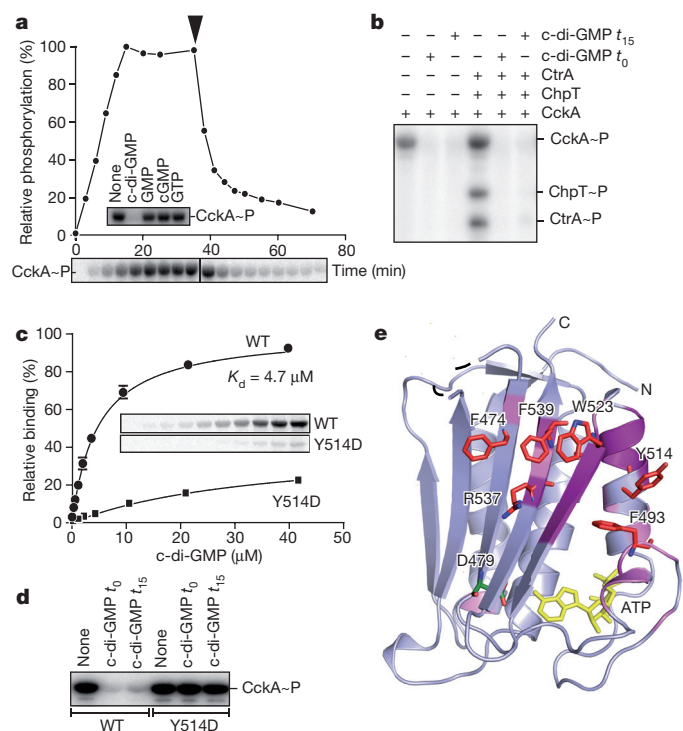
Figure 2 | C-di-GMP binds to the catalytic domain to induce CckA phosphatase activity. **a**, C-di-GMP specifically stimulates CckA dephosphorylation. CckA phosphorylation reactions were started by adding [^{32}P]ATP (0 min) and supplemented with c-di-GMP (75 μM) at the time indicated (arrowhead). The inset shows CckA phosphorylation reactions supplemented with c-di-GMP and other nucleotides (75 μM). Representatives of two technical replicates are shown. **b**, C-di-GMP reverses the phosphate flux of the CckA–ChpT–CtrA phosphorelay. Reactions were run for 30 min and c-di-GMP was added together with [^{32}P]ATP at time 0 min (t_0) or 15 min after reaction start (t_{15}). ~P represents phosphorylation. Representative of three technical replicates is shown. **c**, C-di-GMP binding affinity of CckA. Binding of wild-type CckA and CckA(Y514D) was determined by ultraviolet cross-linking at increasing concentrations of [^{33}P]c-di-GMP (inset) and quantified as shown in the graph. Mean and s.d. were obtained from three technical replicates. **d**, C-di-GMP fails to stimulate phosphatase activity of the CckA(Y514D) mutant. Phosphorylation reactions with wild-type CckA and Y514D mutant protein were analysed without (none) or with c-di-GMP added at 0 min (t_0) or after 15 min (t_{15}). Representatives of three technical replicates are shown. **e**, Homology model of the CA domain of CckA based on a crystal structure of DivL (Protein Data Bank (PDB) ID: 4Q20). Residues that show large ($\Delta\delta(HN) > 2$ s.d.) and intermediate (2 s.d. $> \Delta\delta(HN) > 1$ s.d.) amide chemical shift perturbations upon addition of c-di-GMP are shown in purple and pink, respectively. Side chains of residues that contribute to c-di-GMP binding and c-di-GMP-mediated phosphatase activity are shown in stick representation and coloured in red. A single molecule of ATP (yellow) was modelled into its putative binding site based on homology to CpxA (PDB ID: 4BIX). D479, which is involved in ATP binding, is shown in green. For more information, see the legend of Extended Data Fig. 3.

Agrobacterium tumefaciens (AtCckA). In this organism a spontaneous Y674D substitution in the CA domain of AtCckA was isolated as a motile suppressor of a mutant lacking PleC[16]. We hypothesized that AtCckA is also regulated by c-di-GMP and that in a *pleC* mutant with elevated levels of c-di-GMP (Fig. 1a) the Y674D mutation restores its kinase/phosphatase balance by interfering with c-di-GMP binding. As shown in Extended Data Fig. 2h, autophosphorylation of purified wild-type AtCckA was specifically reversed when c-di-GMP was added, while the AtCckA(Y674D) mutant failed to respond to c-di-GMP. Moreover, c-di-GMP binding to AtCckA(Y674D) was strongly reduced as compared to the wild-type form of the protein (Extended Data Fig. 2h). The equivalent substitution in *Caulobacter* CckA (Y514D) also resulted in strongly diminished c-di-GMP binding

(Fig. 2c). Notably, the CckA(Y514D) mutant showed normal kinase activity but failed to dephosphorylate upon addition of c-di-GMP (Fig. 2d). This was not due to a general lack of phosphatase activity, as the CckA(Y514D) mutant showed an unaltered basal level of phosphatase activity upon ATP depletion (Extended Data Fig. 2i). Together this demonstrates that CckA(Y514D) is compromised for c-di-GMP binding and, as a consequence, cannot switch to the phosphatase mode upon addition of c-di-GMP, resulting in constitutive CckA kinase activity *in vitro*. To define the c-di-GMP binding pocket on the surface of the CA domain we used a combination of structural modelling, biochemical analysis and NMR spectroscopy (Fig. 2e and Extended Data Figs 2i and 3a, b). This approach identified a set of six amino acids, F474, F493, Y514, W523, R537 and F539, which show significant NMR chemical shift perturbations upon c-di-GMP titration experiments and are strictly required for c-di-GMP binding and phosphatase activity but not kinase activity (Fig. 2e and Extended Data Figs 2i and 3b). All of these residues locate in close proximity of Y514 in a homology model of CckA (Fig. 2e). Notably, six of these amino acid residues feature aromatic side chains and are well conserved in CckA homologues (Extended Data Fig. 4). This suggests that c-di-GMP is coordinated by the CA domain of CckA via hydrophobic interactions, akin to the binding mode described for the human STING receptor[17].

Next we set out to test if c-di-GMP executes its important cell cycle role primarily by interfering with the CckA kinase/phosphatase balance *in vivo*. We reasoned that a combination of P*divK*::Tn and *cckA(Y514D)* should cause a similar G1 arrest as observed for a strain lacking c-di-GMP altogether. Moreover, this combination should lead to a cell cycle arrest irrespective of the presence of c-di-GMP (Extended Data Fig. 5a). Indeed, cells carrying P*divK*::Tn and *cckA(Y514D)* showed severe growth defects (Fig. 3a), increased binding of CtrA to the origin region (Extended Data Fig. 5b), and a strong G1 arrest (Fig. 3b and Extended Data Fig. 5c). While this phenotype was independent of PleD, viability of the CckA phosphatase mutant (V366P) strictly depended on c-di-GMP (Fig. 3a, b and Extended Data Fig. 5c). This indicated that downregulation of CckA kinase activity by c-di-GMP is sufficient to balance the kinase/phosphatase activities of the V366P mutant. To corroborate these findings we tested the same *cckA* alleles in strains expressing *divK* from the xylose-dependent promoter P*xyl*. When P*xyl*::*divK* cells were grown in the absence of xylose, DivK dropped below 10% of wild type and, as a consequence, cells developed a mild G1 arrest (Extended Data Fig. 5d). This effect was aggravated in strains expressing *cckA(Y514D)* resulting in a strong

reduction of the DNA/cell mass ratio (Extended Data Fig. 5d), severely reduced growth (Extended Data Fig. 5e), increased CckA phosphorylation levels (Extended Data Fig. 5f) and an overall reduction in the number of chromosomal origins per cell mass (Extended Data Fig. 5g).

Taken together, these experiments lead us to propose a model where two convergent regulatory inputs, DivK and c-di-GMP, control the CckA kinase/phosphatase switch to authorize G1–S transition through the inactivation of the replication initiation inhibitor CtrA (Fig. 1a). Notably, cell-type-specific activity of both PleD and DivK is regulated by DivJ and PleC, two histidine kinase/phosphatase antagonists, which localize to opposite poles of the predivisional cell and during division asymmetrically partition into the daughter cells to determine their respective programs (Fig. 1a)[18]. Thus, the two regulators show similar activation profiles during the cell cycle[7,19], thereby imposing tight coordination between the DivK branch and the c-di-GMP branch of the CckA switch. This connection is further strengthened by the role of DivK as an allosteric activator of the DivJ kinase, a positive feedback mechanism through which both DivK and PleD activity can be rapidly upregulated during G1–S transition[19]. Hence, DivK and c-di-GMP act as molecular connectors between two hierarchical phosphorylation modules, explaining how the cellular dynamics of PleC and DivJ translate into differential activities of the central cell cycle kinase CckA (Fig. 4). Because the parallel morphogenetic program critically depends on PleD activation and the concomitant rise in c-di-GMP concentration[6], c-di-GMP-induced inactivation of CtrA directly couples development to cell cycle progression. This is reminiscent of redundant pathways regulating cell cycle progression in higher eukaryotes, where a multitude of signals converge to control the activity of CDKs[20,21].

In addition to its role in G1–S transition, CckA facilitates cell polarity during division. CckA localizes to both poles of dividing *Caulobacter* cells but adopts differential kinase/phosphatase activities at opposite poles[14,22] (Fig. 1a). The resulting cellular gradient of phosphorylated CtrA was proposed to establish asymmetric replication activities, which propagate to future daughter cells[23]. To test if c-di-GMP contributes to replication asymmetry during division we made use of fluorescent repressor–operator systems to spatially resolve replication initiation events (Extended Data Fig. 6a, b)[24]. While in a majority of wild-type cells chromosome replication originated at the old stalked pole, cells expressing CckA(V366P) or CckA(Y514D) lost replication asymmetry almost entirely (Extended Data Fig. 6c). Cells lacking PleD also partially lost their replication preference for the stalked pole. Because active, phosphorylated PleD specifically localizes to the stalked pole[7,25], we analysed replicative asymmetry in a cdG0 strain expressing a heterologous diguanylate cyclase, DgcZ, from *Escherichia coli*, which is

Figure 3 | C-di-GMP controls CckA activity to initiate chromosome replication. a, The *cckA(Y514D)* allele shows synthetic lethality with P*divK*::Tn. Fivefold serial dilutions of strains containing combinations of *cckA*, P*divK* and *pleD* alleles were incubated on peptone yeast extract plates for 2 days. Representatives of two biological replicates are shown. **b**, Combining the *cckA(Y514D)* and P*divK*::Tn alleles leads to G1 arrest. Exponential cultures of mutants containing combinations of *cckA*, P*divK* and *pleD* alleles were analysed by flow cytometry. Values of DNA content per cell mass are shown relative to wild-type *C. crescentus* and were obtained as described in the legend for Fig. 1c. Mean and s.d. were obtained from three replicates. White and grey bars indicate measurements carried out in wild-type and mutant strains, respectively.

Figure 4 | Model of the regulatory circuitry controlling cell cycle progression in *C. crescentus*. Two intercalated phosphorylation modules control replication initiation through the activity of the replication initiation inhibitor CtrA. When the PleC phosphatase is present at the swarmer pole, PleD and DivK are dephosphorylated. In this situation, phosphorylation modules 1 and 2 are uncoupled and CckA adopts a DivL-imposed kinase mode to activate CtrA and block replication initiation. When the DivJ sensor kinase is present at the stalked pole, phosphorylation module 1 imposes control on module 2. PleD and DivK are phosphorylated (~P), thereby switching CckA into the phosphatase mode and inactivating CtrA. In parallel, c-di-GMP facilitates CtrA degradation via PopA and the ClpXP protease. D and H indicate conserved Asp and His phosphate-acceptor sites.

uniformly distributed in the cell[6,26]. Although expression of DgcZ restored all developmental defects in this strain[6], it failed to establish the characteristic spatial replication bias (Extended Data Fig. 6c). From this we conclude that the spatial organization of c-di-GMP metabolism contributes to cell polarity by differentially regulating CckA at opposite cell poles. For example, a local environment with high levels of c-di-GMP might impose CckA phosphatase activity at the stalked pole. Alternatively, a local trough of c-di-GMP may exist at the swarmer pole with the rest of the cell body containing high levels of c-di-GMP. To distinguish between these possibilities, we made use of a CckA variant that is unable to localize to cell poles because it lacks its membrane anchor (cckAΔTM). Expression of this mutant causes massive over-replication and cell filamentation, arguing that delocalized CckA functions primarily as a phosphatase for CtrA[14]. In agreement with this, expression of cckAΔTM(V366P), lacking phosphatase activity, did not show any adverse effects (Extended Data Fig. 6d). Notably, expression of cckAΔTM(Y514D) in a cdG⁺ strain (Extended Data Fig. 6d) or expression of cckAΔTM in a strain lacking c-di-GMP (Extended Data Fig. 6e) led to a strong G1 arrest, a hallmark of the CckA kinase mode. This indicated that the cellular pool of c-di-GMP strictly imposes phosphatase activity on delocalized CckA molecules.

On the basis of these results we propose that the bulk of the cell volume of dividing *C. crescentus* cells experiences high levels of c-di-GMP and that CckA adopts strong kinase activity at the swarmer pole as a consequence of a microenvironment with low levels of c-di-GMP. This view is consistent with the idea that sequestration of CckA to the swarmer pole creates a microenvironment within the cell where CckA can avoid down-regulation by its other inhibitor, phosphorylated DivK[9,23]. We propose that CckA sequestration to this subcellular site also shields the protein from the cellular pool of c-di-GMP. Ultimately, it is the PleC phosphatase that reduces phosphorylated PleD and DivK levels at this subcellular site and, possibly together with one or several swarmer-pole-specific phosphodiesterases, imposes this spatial regime (Figs. 1a and 4). The input from c-di-GMP might also explain how the entire cellular pool of CckA can be tightly regulated. Throughout the cell cycle, CckA localization is often patchy and dynamic without being strictly limited to polar regions[22]. Since the degree of co-localization of DivK and CckA is unclear, c-di-GMP could effectively maintain CckA in the phosphatase state in all cell types or subcellular regions harbouring high levels the second messenger.

C-di-GMP is only one of several novel nucleotide-based second messengers that have recently been discovered in bacteria[27]. Their global effect on cell physiology raised the question of how these signalling compounds mediate specific cellular responses and how they integrate with other general signalling systems, in particular with two-component phosphorylation networks[28]. Our finding that c-di-GMP acts as a cyclin-like molecule in *C. crescentus* to control the activity of the cell cycle kinase CckA establishes the first direct connection between the two most widespread regulatory networks of bacterial cells. The CckA–ChpT–CtrA pathway is conserved among most known members of the alphaproteobacteria, including important pathogens like *Bartonella* or *Brucella*[29]. This presents the possibility that c-di-GMP-imposed control of sensor histidine kinases might represent a general and widespread regulatory mechanism in bacteria. Considering that c-di-GMP has a major role in regulating virulence and persistence, this provides important new entry points toward a better understanding of the behaviour and propagation of bacterial pathogens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* **15**, 122 (2014).
2. Blanpain, C. & Simons, B. D. Unravelling stem cell dynamics by lineage tracing. *Nature Rev. Mol. Cell Biol.* **14**, 489–502 (2013).
3. Ishidate, T., Elewa, A., Kim, S., Mello, C. C. & Shirayama, M. Divide and differentiate: CDK/Cyclins and the art of development. *Cell Cycle* **13**, 1384–1391 (2014).
4. Morgan, D. O. Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **13**, 261–291 (1997).
5. Römling, U., Galperin, M. Y. & Gomelsky, M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* **77**, 1–52 (2013).
6. Abel, S. *et al.* Bi-modal distribution of the second messenger c-di-GMP controls cell fate and asymmetry during the *Caulobacter* cell cycle. *PLoS Genet.* **9**, e1003744 (2013).
7. Paul, R. *et al.* Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* **18**, 715–727 (2004).
8. Childers, W. S. *et al.* Cell fate regulation governed by a repurposed bacterial histidine kinase. *PLoS Biol.* **12**, e1001979 (2014).
9. Tsokos, C. G., Perchuk, B. S. & Laub, M. T. A dynamic complex of signaling proteins uses polar localization to regulate cell-fate asymmetry in *Caulobacter crescentus*. *Dev. Cell* **20**, 329–341 (2011).
10. Jacobs, C., Domian, I. J., Maddock, J. R. & Shapiro, L. Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division. *Cell* **97**, 111–120 (1999).
11. Biondi, E. G. *et al.* Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* **444**, 899–904 (2006).
12. Quon, K. C., Yang, B., Domian, I. J., Shapiro, L. & Marczynski, G. T. Negative control of bacterial DNA replication by a cell cycle regulatory protein that binds at the chromosome origin. *Proc. Natl Acad. Sci. USA* **95**, 120–125 (1998).
13. Domian, I. J., Quon, K. C. & Shapiro, L. Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* **90**, 415–424 (1997).
14. Chen, Y. E., Tsokos, C. G., Biondi, E. G., Perchuk, B. S. & Laub, M. T. Dynamics of two phosphorelays controlling cell cycle progression in *Caulobacter crescentus*. *J. Bacteriol.* **191**, 7417–7429 (2009).
15. Duerig, A. *et al.* Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev.* **23**, 93–104 (2009).
16. Kim, J., Heindl, J. E. & Fuqua, C. Coordination of division and development influences complex multicellular behavior in *Agrobacterium tumefaciens*. *PLoS ONE* **8**, e56682 (2013).
17. Huang, Y.-H., Liu, X.-Y., Du, X.-X., Jiang, Z.-F. & Su, X.-D. The structural basis for the sensing and binding of cyclic di-GMP by STING. *Nat. Struct. Mol. Biol.* **19**, 728–730 (2012).
18. Wheeler, R. T. & Shapiro, L. Differential localization of two histidine kinases controlling bacterial cell differentiation. *Mol. Cell.* **4**, 683–694 (1999).
19. Paul, R. *et al.* Allosteric regulation of histidine kinases by their cognate response regulator determines cell fate. *Cell* **133**, 452–461 (2008).
20. Choi, Y. J. & Anders, L. Signaling through cyclin D-dependent kinases. *Oncogene* **33**, 1890–1903 (2014).
21. Hochegger, H., Takeda, S. & Hunt, T. Cyclin-dependent kinases and cell-cycle transitions: does one fit all? *Nature Rev. Mol. Cell Biol.* **9**, 910–916 (2008).
22. Angelastro, P. S., Sliusarenko, O. & Jacobs-Wagner, C. Polar localization of the CckA histidine kinase and cell cycle periodicity of the essential master regulator CtrA in *Caulobacter crescentus*. *J. Bacteriol.* **192**, 539–552 (2010).
23. Chen, Y. E. *et al.* Spatial gradient of protein phosphorylation underlies replicative asymmetry in a bacterium. *Proc. Natl Acad. Sci. USA* **108**, 1052–1057 (2011).
24. Robinett, C. C. *et al. In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
25. Paul, R. *et al.* Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *J. Biol. Chem.* **282**, 29170–29177 (2007).
26. Zähringer, F., Lacanna, E., Jenal, U., Schirmer, T. & Boehm, A. Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure* **21**, 1149–1157 (2013).
27. Danilchanka, O. & Mekalanos, J. J. Cyclic dinucleotides and the innate immune response. *Cell* **154**, 962–970 (2013).
28. Capra, E. J. & Laub, M. T. Evolution of two-component signal transduction systems. *Annu. Rev. Microbiol.* **66**, 325–347 (2012).
29. Brilli, M. *et al.* The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.* **4**, 52 (2010).

Supplementary Information is available in the online version of the paper.

Author Contributions C.L., S.O., S.A., S.S. and U.J. initiated the project. All authors designed experiments. S.O. carried out genetic experiments. C.L. and S.S. carried out biochemical experiments. C.L. and S.O. carried out microscopy experiments. R.B. and S.H. carried out NMR experiments. B.N.D. and T.S. contributed to structural analysis. U.J., S.O. and C.L. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to U.J. (urs.jenal@unibas.ch).

METHODS

Strains and plasmids. Strains used in this study are listed in Supplementary Table 1. *Caulobacter* strains are grown in peptone yeast extract medium (PYE) or minimal medium supplemented with glucose (M2G) at 30 °C[30]. When necessary, medium was supplemented with glucose (0.2%), xylose (0.03% or 0.3% as indicated), and antibiotics as described[31]. When synchronized *Caulobacter* cell cultures were used, newborn cells were harvested by LUDOX density-gradient centrifugation method[6]. Generalized CR30 phage transduction was performed as described[30].

Strain UJ6777 was constructed by sequential two-step transduction using a CR30 phage lysate of MT15 (ref. 32). Strains UJ8306, UJ8307, UJ8308 and UJ8314 were constructed by sequential two-step transduction using a CR30 phage lysate of MT16 (ref. 32). UJ8312 was constructed by sequential transduction using parental strain SoA1273 and CR30 phage lysates of MT16 and UJ6777. Strains UJ7212 and UJ7214 were constructed using a parental NA1000 strain and suicide vectors pNTPS-CckA(Y514D) and pNPTS-CckA(V366P), respectively. To construct strains UJ6861 and UJ7304, *lacA*::Ω was transduced into UJ5065 using a phage lysate of UJ6168. Subsequently, pBlue-*pleD* was transformed into the transductant, yielding UJ6861. *xylX*::*tipNgfp* was transformed into UJ6861, yielding UJ7304. Strain UJ7525 was constructed by sequential two-step transduction. First, *xylX*::pPA28 was transduced into NA1000 using a phage lysate of UJ286. Resulting kanamycin-resistant colonies were subsequently transduced with Δ*divK*::Ω using a phage lysate of strain CJ403. Strains UJ7527, UJ7529, UJ7618, UJ7619 and UJ7620 were constructed similarly using UJ7212, UJ7214, UJ7417, UJ7418 and UJ7419, respectively, as a parental strain instead of NA1000. Strains UJ7417, UJ7418 and UJ7419 were constructed by double homologous recombination using pNPTS-CckA-(3×Flag) and parental strain NA1000, UJ7212 or UJ7214. Strains UJ7511 and UJ7512 were generated by integration of pMCS1-CckA into NA1000 and UJ7212. Strains UJ7873 and UJ7992 were constructed by double homologous recombination using pNPTS-XdivK and a parental strain NA1000 (for UJ7873) or UJ5065 (for UJ7992), respectively. Strains UJ7939 and UJ7940 were constructed by transformation of pMCS5-k2t into NA1000 P*divK*::Tn and NA1000 Δ*pleD* P*divK*::Tn, respectively.

Plasmids and oligonucleotides used in this study are listed in Supplementary Tables 2 and 3, respectively. To construct pBlue-*pleD*, a 2.5 kb fragment containing the *pleD* gene under control of the *divK* promoter was amplified using pPA41 and primers 5156 and 104, followed by digestion with ScaI and ligation to the ScaI fragment of pJC389. The insert was verified by DNA sequencing. To construct pNPTS-XdivK, the upstream (694 bp) of *divK* was amplified using NA1000 genome and primers 6151 and 6152. The product was digested with XhoI and SacII, ligated to the SalI-SacII fragment of pXMCS-1, resulting in pXdivKr. In parallel, a 1 kb fragment containing the *divK* gene and its downstream region was amplified using SH119 genome and primers 6153 and 6154, followed by digestion with EcoRI and NdeI and ligation to the EcoRI-NdeI fragment of pXdivKr, resulting in pXdivKrl. Finally, the EcoRI-SphI fragment of pXdivKrl was subcloned into pNPTS138, yielding pNPTS-XdivK. For pMCS5-k2t, a part (444 bp) of the *nptII* gene was amplified by PCR using pAlmar1 and primers 6693 and 6694, followed by digestion with KpnI and SacI and ligation to the KpnI-SacI fragment of pMCS-5. For pXTCYC4-tipNgfp, the *tipNgfp* gene was amplified using pMR20-tipNgfp (UJ6350) and primers 5242 and 105, followed by ligation into pGEM vector. The resulting plasmid was digested with NdeI and SacI, followed by ligation to the NdeI-SacI fragment of pXTCYC-4. To construct pET28a-His-MBP, the His-MBP fragment was amplified using pHIS-MBP-DEST and CckA and primers 5196 and 5278 followed by digestion with BamHI and NcoI and subsequent ligation into pET28a. To construct pET-CckA, CckA was amplified using primers 5276 and 5277 followed by digestion with BamHI and SalI and ligation into pET28a-His-MBP. To introduce point mutations into pET-CckA SOE-PCR was used. Generally pET-CckA was used as template with 5276/5277 as outside primers and internal mutagenic primers to introduce mutations. The following internal primers were used to introduce point mutations: V366P (5134/5135), Y514D (5448/5449), F474A (7725/7726), D479A (7727/7728), F493A (7729/7730), W523A (7735/7736), R537A (5502/5503), F539A (7737/7738). After fusion PCR inserts were BamHI and SalI digested and ligated into pET28a-His-MBP.

To amplify truncated *cckA* fragments the following primers were used. pET-CckA S72–I573 (5276/5454), pET-CckA G571–A691 (5455/5277), pET-CckA P541–A691 (5456/5277), pET-CckA F496–A691 (5457/5277), pET-CckA V417–A691 (5458/5277), pET-CckA S72–P546 (5276/5644), pET-CckA A312–A691 (5646/5277), pET-CckA I292–P546 (5280/5644), pET-CckA A371–A691 (5645/5277). Resulting PCR products were digested with BamHI and SalI and ligated into pET28a-His-MBP. To generate pET-CckA V417–A691 (N-ZIP) and pET-CckA A371–A691 (N-ZIP) N-ZIP was amplified using primers 5647 and 5648 using pUT18C-zip as template. The resulting PCR product was digested with BamHI and ligated into pET-CckA V417–A691 and pET-CckA A371–A691.

Correct orientation of insert verified by sequencing. pET21b-CckA Q379–A545 was generated using primers 7244 and 7249 and chromosomal DNA as template. The resulting PCR product was NotI and NdeI digested and ligated into pET21b.

To construct pET-AgroCcka, *cckA* was amplified from chromosomal DNA of *Agrobacterium tumefaciens* C58 using primers 6430 and 6431. To introduce Y674D mutation SOE-PCR was used using mutagenic primers (6436/6437). Inserts were digested with BamHI and SalI and ligated into pET28a-His-MBP.

To construct pNPTS-CckA(V366P) the insert of pET-CckA(V366P) was cut out using BamHI and HindIII and ligated into pNPTS138. To construct pNPTS-CckA(Y514D) a fragment was amplified using primers 5458 and 670 and pET-CckA(Y514D) as a template. PCR product was BamHI and HindIII digested and ligated into pNPTS138. To construct pNPTS-CckA-(3×Flag) a first fragment was amplified using primers 5456 and 5938 and *cckA* as template. A second PCR was run on the first PCR product to extend 3×Flag using primers 5456 and 5939. Downstream fragment was amplified from chromosomal NA1000 DNA using primers 5940 and 5941. The downstream and 3×Flag fragments were fused by SOE-PCR using primers 5456 and 5941. This PCR product was BamHI and HindIII digested and ligated into pNPTS138.

To construct pSA241.1 *ctrA* was amplified from chromosomal DNA using primers 1505 and 3708. PCR product was digested with BamHI and HindIII and ligated into pTRcHisA. pTRc-CtrA(D51E) was generated as pSA241.1 except that SOE-PCR was used to introduce D51E mutation using mutagenic primers 4818 and 4819 and outside primers 1505 and 3708. To construct pMCS-CckA a fragment from pET-CckA P541–A691 was cut out using SacI and HindIII and subcloned into pMCS-1.

To construct pBXMCS-CckA, pBXMCS-CckA(V366P) and pBXMCS-CckA(Y514D) a fragment was amplified from plasmid DNA (pET-CckA, pET-CckA(V366P) or pET-CckA(Y514D)) using primers 7200 and 7369. The resulting PCR product was NdeI and EcoRI digested and ligated into pBXMCS-2.

Screen for synthetic lethality. Although a mutant unable to synthesize c-di-GMP (cdG0) shows pronounced abnormalities in cellular DNA content and cell morphology, its overall growth and viability were not affected. We reasoned that c-di-GMP, together with redundant pathways, maintains core bacterial cell cycle processes like cell division or chromosome replication. To probe for such interactions a genetic screen for synthetic lethality was adapted for the *C. crescentus* cdG0 strain[33]. The screen identifies transposon (Tn) mutants that stably retain the plasmid (pBlue-pleD) under non-selective conditions. pBlue-pleD expresses the diguanylate cyclase PleD as the sole source of c-di-GMP present in the screening strain. While this plasmid is unstable and easily lost in the absence of selection, a transposon insertion generating synthetic lethality in the cdG0 strain should render pBlue-pleD essential for growth, which in turn results in its stable maintenance without antibiotic selection. pBlue-pleD also carries *lacA*, a gene required for the metabolism of β-galactosides[34]. Because the *lacA* copy was deleted in the screening strain, its ability to metabolize β-galactosides relies on plasmid-borne *lacA*. Consequently, mutants that stably maintain the plasmid yield solid blue colonies on plates supplemented with X-gal, while mutants that loose the plasmid are easily recognized by their blue/white sectored appearance.

Random transposon mutagenesis was performed by transforming a transposon donor plasmid pAlmar1 into the screening strain (UJ7304 or UJ6861). Total 142,000 transposon mutants were grown at 30 °C for 1 week on PYE plates supplemented with 20 μg ml kanamycin and 40 μg ml X-gal. Solid blue colonies were streaked and incubated on an X-gal supplemented PYE plate for 3 days at 30 °C. Transposons were mapped by a two-step arbitrary PCR as described[35]. In brief, downstream of the transposon was amplified by first PCR using primers 1228 and 1365. After purification, DNA fragments were further enriched by second PCR using primers 1365 and 1657. The products were sequenced using the primer 2614 (S.O. and U.J., unpublished result).

Spot growth assay. The cell density of each culture was adjusted to attenuance (*D*) at 660 nm of 0.014, followed by preparation of serial fivefold dilutions. Five microlitres of the culture was spotted onto PYE plate containing appropriate supplements and incubated for 2 days at 30 °C.

Western blotting. Anti-PleD (1:2,000), anti-DivK (1:5,000), and anti-Flag (1:10,000) antibodies were used as primary antibodies, which were detected by HPR-conjugated rabbit anti-mouse or swine anti-rabbit secondary antibodies (Dako), followed by development with ECL detection reagents.

Flow cytometry. This assay was performed essentially as described[15]. In brief, exponentially growing cells (100 μl) were fixed in ice-cold 70% ethanol. For rifampicin treatment, cells were incubated for 1 h at 30 °C in the presence of rifampicin (final 30 μg ml) before fixation. Cells were harvested by centrifugation, re-suspended in 0.5 ml FACS buffer (10 mM Tris HCl pH 7.5, 1 mM EDTA, 50 mM sodium citrate, and 0.01% TritonX-100) containing 0.1 mg ml RNaseA, and incubated at room temperature for 30 min. After harvesting cells by centrifugation,

DNA was stained in 1 ml FACS buffer including 1.5 µM YO-PRO-1 iodide (Invitrogen) at room temperature for 2 h. The fluorescent intensity and the light scattering were analysed using FACS Canto II (BD Biosciences).

Protein purification. Expression plasmids were transformed into *E. coli* BL21 (DE3) cells. Cells were grown in liquid broth at 30 °C to an OD_{578} of 0.5 and subsequently induced with 300 µM IPTG for 4 h. Cells were pelleted and frozen at −80 °C. Proteins were purified on a Äkta purifier using 1 ml HisTrap HP columns (GE Healthcare) and if higher purity was desired proteins were run on a size-exclusion column (HiLoad 16/60 Superdex 200). For the purification the following buffers were used: lysis buffer (wash buffer supplemented with protease inhibitor), wash buffer (16 mM Na_2HPO_4, 3.6 mM KH_2PO_4, 5.4 mM KCl, 500 mM NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, pH 7.0), elution buffer (16 mM Na_2HPO_4, 3.6 mM KH_2PO_4, 5.4 mM KCl, 500 mM NaCl, 2 mM β-mercaptoethanol, 500 mM imidazole, pH 7.0), storage and activity buffer (10 mM HEPES-KOH, 50 mM KCl, 10% glycerol, 0.1 mM EDTA, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, pH 8.0). The *A. tumefaciens* CckA homologue was stored in an optimized buffer (10 mM HEPES, 125 mM KAc, 10% glycerol, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, pH 7.5).

Kinase and phosphatase assays. Generally, kinase and phosphatase assays were adapted from refs 14 and 36. Reactions were run in activity buffer in presence of 500 µM ATP and 5 µCi [γ^{32}P]ATP (3,000 Ci mmol$^{−1}$, Hartmann Analytic) at room temperature. Additional nucleotides were added at indicated time points. Reactions were stopped with SDS sample buffer and subsequently loaded (or stored on ice) on 10% SDS gels. Wet gels were exposed to phosphor screen (0.5–3 h) before being scanned using a Typhoon FLA 7000 imaging system (GE Healthcare). When needed, ATP was depleted from the reaction mixtures by the addition of 1.5 U hexokinase (Roche) and 5 mM D-glucose after 15 min of phosphorylation.

Ultraviolet cross-linking with [^{33}P]c-di-GMP. ^{33}P-labelled c-di-GMP was produced using [α^{33}P]GTP (Perkin Elmer) and the *E. coli* diguanylate cyclase DgcZ. Purification of DgcZ and production of c-di-GMP was performed as previously described[26]. Purified proteins were incubated with labelled c-di-GMP for 30 min in activity buffer at room temperature. Proteins were cross-linked at 254 nm for 3 min at 4 °C and then diluted into SDS sample buffer as described[37]. After 5 min of boiling the samples were separated by SDS–PAGE. Gels were dried and exposed to a phosphor screen overnight and then scanned on an imaging system. Band intensities were quantified using ImageJ and binding curves were fitted with GraphPad Prism.

Quantification of CckA~P levels *in vivo*. Colonies grown on PYE plates or PYE supplemented with 0.2% glucose were re-suspended in PYE and adjusted to the same OD. Cells were pelleted and re-suspended in 100 µl lysis buffer (10 mM Tris-HCl, 4% SDS, one tablet phos-stop (Roche), pH 7.5). Lysates were diluted into SDS sample buffer and analysed by SDS–PAGE gels (7.5%) supplemented with 50 mM phos-tag acrylamide (Wako) and 100 mM manganese chloride. Gels were run at 4 °C at 100 V for 3 h. Before immunoblotting the gels were incubated for 10 min in transfer buffer (1× TrisGlycine, 20% ethanol) containing 1 mM EDTA and for another 10 min in transfer buffer without EDTA. Proteins were detected using anti-flag antibodies.

NMR experiments. NMR spectra were recorded at 20 °C on Bruker Avance-700 and -900 spectrometers equipped with cryogenic triple-resonance probes. CckA-CA protein samples were prepared in 30 mM Tris-HCl at pH 7.5 with 100 mM NaCl, 5 mM $MgCl_2$ and 2 mM DTT in 95%/5% H_2O/D_2O. For the sequence-specific backbone resonance assignment of [U-^{2}H, ^{15}N, ^{13}C]CckA-CA, the following NMR experiments were recorded: 2D [^{15}N,^{1}H]-TROSY-HSQC[38], 3D TROSY-HNCA[39] and 3D [^{1}H,^{1}H]NOESY-^{15}N-TROSY with a NOE mixing time of 100 ms[40]. For the c-di-GMP binding experiments a series of 2D [^{15}N,^{1}H]TROSY-HSQC spectra of 380 µM [U-^{15}N]CckA-CA were recorded with c-di-GMP concentrations of 0 mM, 0.04 mM, 0.24 mM, 0.6 mM, 1.32 mM, 2.54 mM and 4.27 mM. Chemical shift perturbations ($\Delta\delta(HN)$) of amide moieties were calculated as:

$$\Delta\delta(HN) = \sqrt{\frac{(\delta H_{ref} - \delta H)^2 + \left(\frac{\delta N_{ref} - \delta N}{5}\right)^2}{2}}$$

ChIP and qPCR. ChIP was performed as described previously[41]. Cells were grown in 20 ml PYE until $D_{660\,nm}$ reached ∼0.2. To cross-link protein–DNA complexes, 0.2 ml of 1 M Na-phosphate pH 7.6 (final 10 mM) and 540 µl of 37% formaldehyde (final 1%) were added to the culture and incubated at room temperature for 10 min, followed by incubation on ice for 30 min. After harvesting cells by centrifugation (2,600*g* for 30 min), cells were washed twice in 20 ml of ice-cold PBS and re-suspended in buffer A (10 mM Tris HCl pH 8.1, 20% sucrose, 50 mM sodium chloride, 10 mM EDTA, and 20 mg ml$^{−1}$ lysozyme) to adjust final $D_{660\,nm}$ of 8. After incubation at 37 °C for 30 min, the equal volume of buffer IP (100 mM Tris HCl pH 7.5, 300 mM sodium chloride, 2% Triton X-100, and 2× complete

protease inhibitor (Roche)) was added to the sample and incubation was continued for 15 min at 37 °C. Subsequently genomic DNA was sheared by sonication and cell debris was removed by centrifugation, yielding a clear lysate. A portion (30 µl) of the lysate was mixed with 70 µl TE buffer including 1% SDS to use as an input DNA. Another portion (0.2 ml) was incubated with the anti-CtrA antibody (0.8 µl) in a cold room for >12 h with gentle agitation. The incubation was continued for 4 h in the presence of Protein-A agarose beads (100 µl slurry), followed by washing beads seven times in 0.5 ml of ice-cold buffer IP (50 mM Tris HCl pH 7.5, 150 mM sodium chloride, and 1% Triton X-100) and two times in 0.5 ml of TE buffer. The beads were re-suspended in 0.1 mL TE buffer containing 1% SDS and incubated at 65 °C for >12 h to reverse cross-linking. DNA samples were purified using a DNA purification kit (Macherey-Nagel). qPCR was performed as described previously[42]. In brief, DNA of the origin region was amplified by quantitative PCR StepOne Plus (Applied Biosystems) using Power SYBR Green PCR Master Mix (Applied Biosystems) and primers 6708 and 6709. To measure background signals, a part of the *ctrA* coding region was amplified similarly using primers 6710 and 6711.

Scoring replication initiation using an origin-specific fluorescent repressor–operator system. To visualize DNA replication in individual cells we made use of a fluorescent repressor–operator system that allows tracking of chromosomal origins during the cell cycle[24,32]. Cells used in these experiments produce TetR–YFP and harbour tet operator (*tetO*) sites near the origin of replication. Binding of TetR–YFP molecules to the operator arrays yields a fluorescent signal that stains the origin. Upon initiation of chromosome replication, duplicated operator arrays produce two discrete fluorescent foci. This enables tracking of replication initiation events at opposite poles of the predivisional cell. A (*tetO*)$_n$ cassette was integrated at the *cc0006* locus by phage transduction using a phage lysate of MT16. TetR–YFP was expressed from the xylose promoter by induction with 0.3% xylose. *tetR-yfp* was introduced by transduction using a phage lysate of MT15.

Replication asymmetry in the predivisional cell was investigated as described previously[23]. In brief, cells grown overnight in PYE were diluted into PYE and grown to OD of 0.3. One hour before microscopy cells were induced with 0.3% xylose and subsequently mounted on an agar-pad supplemented with PYE 0.3% xylose and cephalexin (10 µg ml$^{−1}$), followed by time-lapse microscopy with 10 min intervals for 5 h.

Microscopy. Differential interference contrast (DIC), phase-contrast, and fluorescent microscopy analyses were performed using a DeltaVision system, Olympus IX71 microscope, and Photometrix CoolSnap HQ2 camera. Cells were mounted on 1.2% agar containing appropriate supplements. For statistics, cell length and the number of fluorescent foci were analysed using MicrobeTracker (http://microbetracker.org).

Statistical analysis. For biochemistry we performed experiments as described earlier. All results documented are highly reproducible. Where indicated, mean values and standard deviations were obtained from at least three independent experiments (biological replicates). For flow cytometry, we analysed more than two biological replicates. The results were highly reproducible with reasonable standard deviation. For replicative asymmetry measurements, we used *z*-test to show that confidence levels for all measurement were above 99.9% (http://www.mccallum-layton.co.uk/tools/statistic-calculators). No statistical method was used to predetermine sample size.

30. Ely, B. Genetics of *Caulobacter crescentus*. *Methods Enzymol.* **204,** 372–384 (1991).
31. Thanbichler, M., Iniesta, A. A. & Shapiro, L. A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in *Caulobacter crescentus*. *Nucleic Acids Res.* **35,** e137 (2007).
32. Viollier, P. H. *et al.* Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl Acad. Sci. USA* **101,** 9257–9262 (2004).
33. Bernhardt, T. G. & de Boer, P. A. J. Screening for synthetic lethal mutants in *Escherichia coli* and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity. *Mol. Microbiol.* **52,** 1255–1269 (2004).
34. Arellano, B. H., Ortiz, J. D., Manzano, J. & Chen, J. C. Identification of a dehydrogenase required for lactose metabolism in *Caulobacter crescentus*. *Appl. Environ. Microbiol.* **76,** 3004–3014 (2010).
35. Navarre, W. W. *et al.* PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. *Mol. Cell* **39,** 209–221 (2010).
36. Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G. & Laub, M. T. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol.* **3,** e334 (2005).
37. Christen, M. *et al.* DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. *Proc. Natl Acad. Sci. USA* **104,** 4112–4117 (2007).
38. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA* **94,** 12366–12371 (1997).

39. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA* **95,** 13585–13590 (1998).

40. Zuiderweg, E. R. & Fesik, S. W. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28,** 2387–2391 (1989).

41. Quisel, J. D., Lin, D. C. & Grossman, A. D. Control of development by altered localization of a transcription factor in *B. subtilis. Mol. Cell* **4,** 665–672 (1999).

42. Taylor, J. A., Ouimet, M.-C., Wargachuk, R. & Marczynski, G. T. The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol. Microbiol.* **82,** 312–326 (2011).

43. Hung, D. Y. & Shapiro, L. A signal transduction protein cues proteolytic events critical to *Caulobacter* cell cycle progression. *Proc. Natl Acad. Sci. USA* **99,** 13160–13165 (2002).

44. Kjaergaard, M. & Poulsen, F. M. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **50,** 157–165 (2011).

45. Hildebrand, A., Remmert, M., Biegert, A. & Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins* **77** (Suppl. 9), 128–132 (2009).

46. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
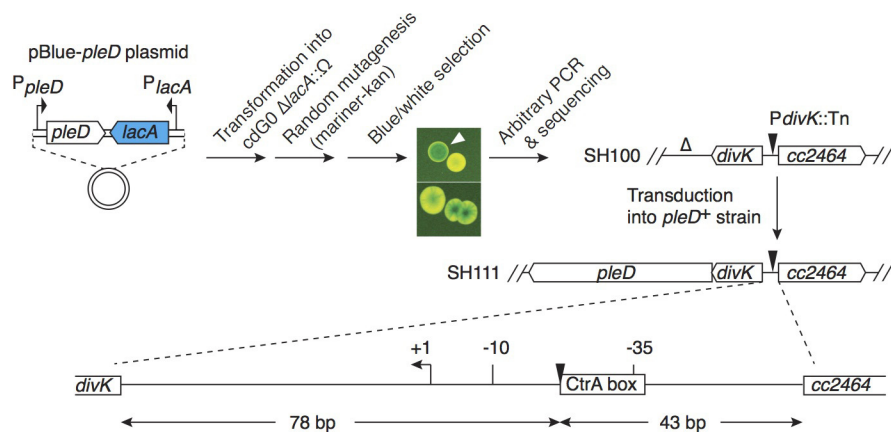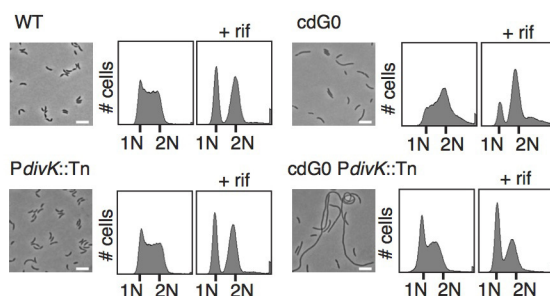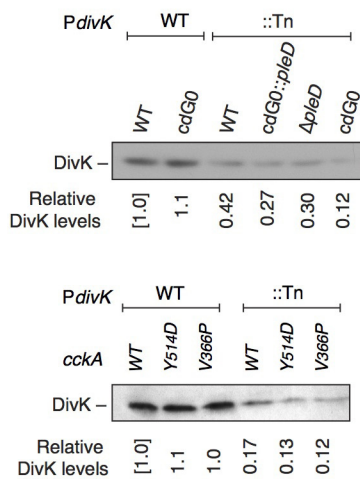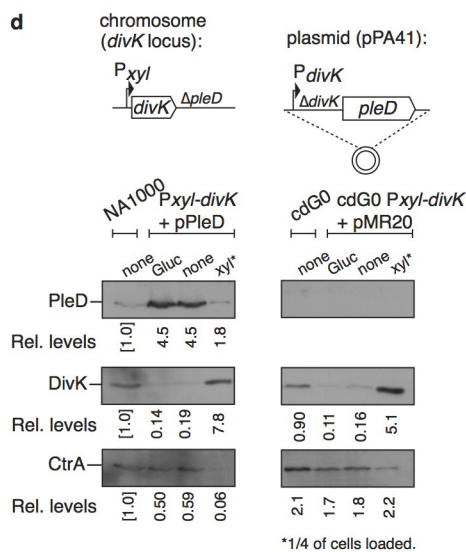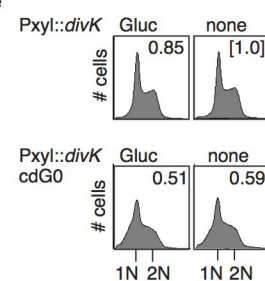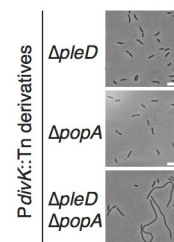
Extended Data Figure 1 | Characterization of P*divK*::Tn and P*xyl*::*divK* derivatives. a, Schematic of the synthetic lethality screen. pBlue-*pleD* is a low copy number plasmid carrying both *pleD* and *lacA* genes, each with its own promoter. The *lacA* gene encodes a subunit of the LacABC dehydrogenase responsible for the breakdown of β-galactosides in *C. crescentus*[34]. An open arrowhead in the top panel indicates a representative blue colony on an X-gal agar plate. Out of 142,000 independent transformants, representative white colonies or colonies with blue sectors indicating segregation of the unstable pBlue-pleD plasmid are indicated in the upper and lower panel. Genomic organizations of the *divK* locus in strains SH100 and SH111 are shown schematically. The exact position of the transposon insertion (P*divK*::Tn) in the *divK* promoter region adjacent to the CtrA box is indicated by closed arrowheads. The transcription start site (+1) and the −10 and −35 elements are shown[43]. Mapping of the transposon to the CtrA-binding site in the *divK* promoter region might imply that this lesion reduces *divK* expression by interfering with CtrA-mediated positive control[11]. **b**, Cell morphology and chromosome replication activity. Indicated strains were analysed microscopically and by flow cytometry to measure DNA content. Cells were grown with or without rifampicin (rif) as indicated. Chromosome equivalents (N) are indicated. Phase-contrast images are shown with scale bars of 5 μm.

Representatives of two biological replicates are shown. **c**, DivK levels deduced by immunoblot analysis. Cells grown in peptone yeast extract were harvested at an OD_{660} of ~0.2 and subjected to SDS-13% PAGE, followed by immunoblot analysis using anti-DivK antibodies. The intensities of the DivK bands were quantified using ImageJ and are shown as relative values to NA1000 wild-type levels. Representative of two biological replicates is shown. **d**, Subcellular levels of PleD, DivK and CtrA in the P*xyl*::*divK* derivatives. Cells of strains NA1000, UJ5065, UJ8012 and UJ8013 grown in peptone yeast extract (none) or peptone yeast extract supplemented with 0.2% glucose or 0.03% xylose were analysed by immunoblots. The intensities of the protein bands were quantified using ImageJ and are shown as relative values to wild-type NA1000. The vector control (pMR20) is indicated. Representatives of two biological replicates are shown. **e**, Chromosome replication activity of wild-type (UJ8012) and cdG0 (UJ8013) strains expressing *divK* from the P*xyl* promoter. Strains were grown exponentially in peptone yeast extract (none) or peptone yeast extract supplemented with glucose (Gluc), followed by flow cytometry analysis. Representatives of two biological replicates are shown. **f**, Effect of P*divK*::Tn on cell morphology in strains lacking *pleD*, *popA*, or both. Scale bar, 5 μm. Representative of two biological replicates is shown.
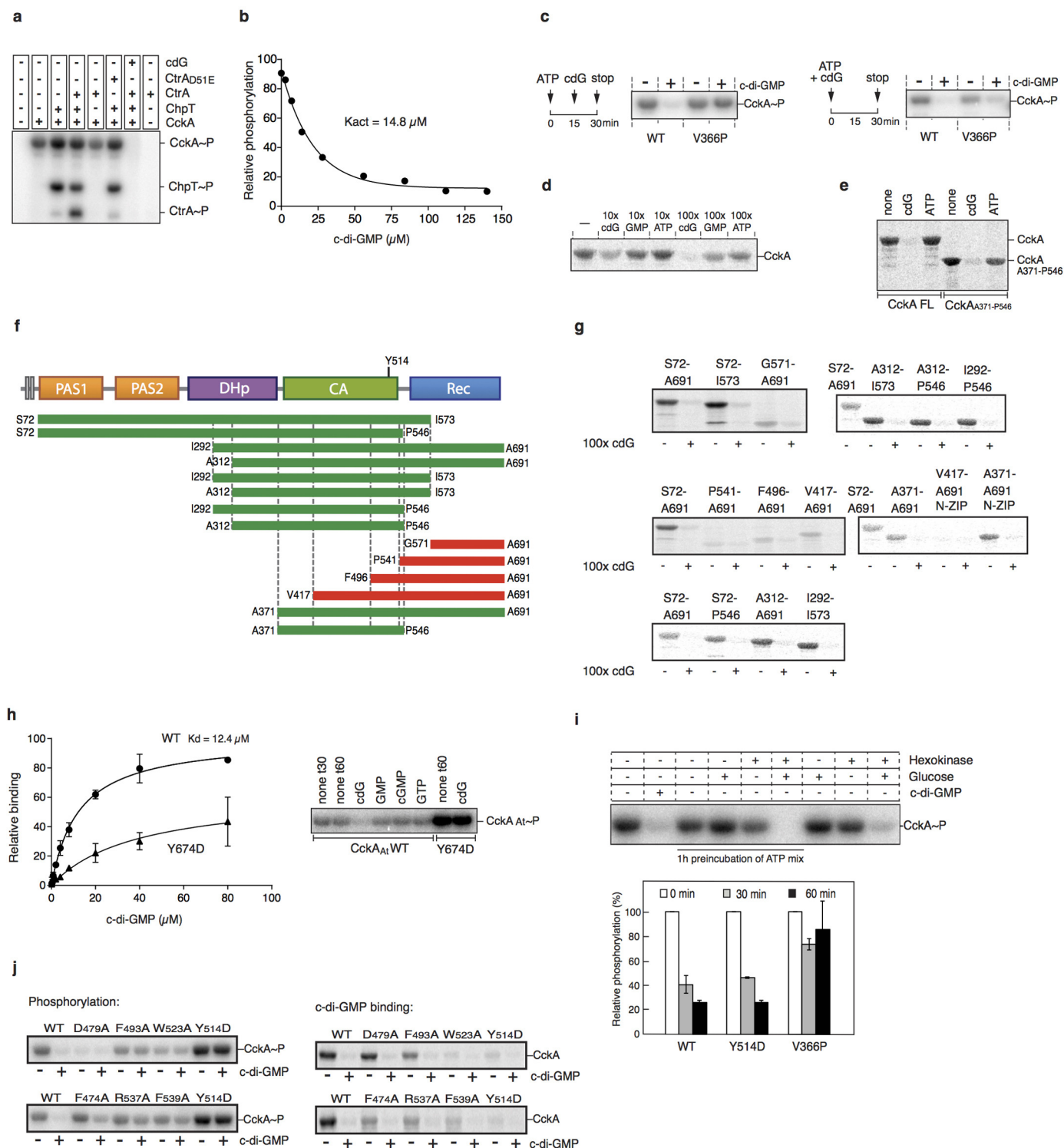
Extended Data Figure 2 | C-di-GMP binds to CckA to induce phosphatase activity. **a**, C-di-GMP inhibits the CckA phosphorelay. *In vitro* phosphorylation reactions with purified proteins (+) in the presence or absence of c-di-GMP (75 μM). A CtrA mutant (D51E) lacking the phosphate-acceptor site is shown as a control. Phosphorylated proteins are marked ~P. The weak band with a size similar to CtrA (lines 3 and 6) corresponds to a phosphorylated breakdown product of ChpT. Representatives of three technical replicates are shown. **b**, C-di-GMP stimulates CckA dephosphorylation. Phosphorylation reactions with purified CckA were carried out as outlined in Fig. 2a. After reaching saturation, dephosphorylation was initiated by the addition of increasing concentrations of c-di-GMP. Reactions were run for 15 min and were analysed by autoradiography. Representative results of two technical replicates are shown. **c**, C-di-GMP inhibits CckA autophosphorylation. Purified wild-type CckA and phosphatase mutant (V366P) were incubated with [^{32}P]ATP and with (+) or without (−) c-di-GMP (75 μM) as indicated. C-di-GMP was added at the time points indicated. Representatives of three technical replicates are shown. **d**, C-di-GMP specifically binds to CckA. Purified CckA protein was incubated with ^{33}P-labelled c-di-GMP and cross-linked with ultraviolet light in the presence or absence of a tenfold or 100-fold excess of competing non-labelled nucleotides as indicated. Representatives of three technical replicates are shown. **e**, The CA domain of CckA specifically binds c-di-GMP. Purified full length CckA (FL, lacking N-terminal transmembrane domains) and the minimal binding unit (see **f**) was incubated with ^{33}P-labelled c-di-GMP and cross-linked with ultraviolet light in the presence or absence of a 100-fold excess of non-labelled ATP or c-di-GMP as indicated. Representatives of three technical replicates are shown. **f**, Schematic of the domain architecture and truncated constructs of CckA. Amino acids marking the boundaries of each construct are indicated. Constructs marked by green and red bars showed c-di-GMP binding or failed

to bind c-di-GMP, respectively. **g**, Truncated versions of the CckA proteins indicated in **a** were expressed, purified and analysed for c-di-GMP binding using ultraviolet cross-linking of ^{33}P-labelled c-di-GMP[37] (10 μM) in the presence (+) or absence (−) of a 100-fold excess of non-labelled c-di-GMP (1 mM). Samples were analysed by SDS–PAGE and autoradiography as indicated. Representatives of two technical replicates are shown. **h**, Left, AtCckA, the CckA homologue of *A. tumefaciens* binds c-di-GMP. The c-di-GMP binding affinities of wild-type AtCckA and the AtCckA(Y674D) mutant protein were determined by ultraviolet cross-linking at increasing concentrations of [^{33}P]c-di-GMP. Relative binding units and affinities are shown. Error bars are standard deviations. Averages and standard deviations were obtained from three technical replicates. Right, AtCckA is regulated by c-di-GMP. Wild-type AtCckA and AtCckA(Y674D) mutant were incubated with [^{32}P]ATP (0 min) and supplemented with c-di-GMP and other nucleotides (75 μM) at 30 min. Fractions were removed after 30 min or 60 min as indicated and analysed by autoradiography. Representative of two technical replicates is shown. **i**, Phosphatase activity of CckA alleles in the absence of ATP. Reactions were allowed to autophosphorylate for 15 min before hexokinase and D-glucose were added to rapidly deplete ATP. A representative gel image for wild-type CckA is shown (top). Kinetic analysis revealed that CckA(Y514D) retains wild-type-like phosphatase activity (bottom). Error bars are standard deviations. Averages and standard deviations were obtained from three technical replicates. **j**, Mutational analysis of amino acids contributing to c-di-GMP binding and phosphatase control. Purified CckA wild-type and mutant forms were analysed for phosphorylation activity and [^{33}P]c-di-GMP binding as indicated above. Note that the residue D479 is involved in ATP binding. Consequently, the D479A mutant lacks kinase activity, but is unaltered in its ability to bind c-di-GMP. Representatives of two technical replicates are shown.

a

G490 G522
G386 G515
G446
G445 G464
G481 G509 G444 G470
G483 G533
V417 T456 G503 G411 V439
D487 S422 K443 T452
Q416 D458 D521 S520
N531 T408 D479 V428 R412 T426 R401
M429 V433 R397 D480 S390 I389 L463 A442 V394 R398
D403 R436 F493 N434 A441 V404
D413 D437 F496 Q519 E532 L399 K491 I492
E393 L392 I461 L400 F497 L431 V488 L395
R454 W523 L396 E425 V382 L388 K518 S528 M489 E459
L540 D420 D469 L510 E550 I484 A534 E391
R418 D471 I450 A460 V517 L508 A427 A548 D494
R537 I538 R457 D384 R380 A547 A546 I516 A435 V513
V447 F536 F474 Q423 E387 A438 I524
F539 A453 Y410 A432 D409 K405 H556 V526
H525 E476 V448 I475 A545 H527
R449 R451 L385 R529
A473 L383 K421 I407 Y543 E544
V477

δ₂(¹H) (p.p.m.)

δ₁(¹⁵N) (p.p.m.)

CckA-CA residue number

Δδ(¹³Cα) (p.p.m.)

β1 α1 β2 β3 α2 β4 α3 β5 α4 α5 β6 β7

```
                379                          400                  420
CckA  QREVLDLGELISEFEVLLRRLLRE-DVKLITDYGRDLPQVRADKSQL
DivL  EIEDIRVSDLLLNAQERALKDAQLGGVTLAVECEEDVGLIRGDGKRL
CckA      β1      α1         β2      β3   α2
DivL

                        440              460
CckA  ETAVMNLAVNARDAVRAAKGGGVVRIRTARLTRDEAIQLGFPAADGD
DivL  AQTLDHLVENALRQTPP---GGRVTLSARR-------------ALG
CckA        α2              β4      α3
DivL

          480                500
CckA  TAFIEVSDDGPGIPPDVMGKIFDPFFTTKPVGEGTGLGLATVYGIVK
DivL  EVRLDVSDTGRGVPFHVQAHIFDRFVGRDR--GGPGLGLALVKALVE
CckA     β5         α4                      α5
DivL

          520              540
CckA  QSDGWIHVHSRPNEGAAFRIFLPVYEA
DivL  LHGGWVALESEPGNGSTFTCHLPETQQ
CckA      β6      β7
DivL
```

b

Δδ(HN) (p.p.m.)

CckA-CA residue number

0.38 mM Cck-CA
0.36 mM Cck-CA + 1.3 mM c-di-GMP
0.35 mM Cck-CA + 4.3 mM c-di-GMP

V447 F539
H525 R449
L445

δ₁(¹⁵N) (p.p.m.)

δ₂(¹H) (p.p.m.)

Extended Data Figure 3 | Characterization of the c-di-GMP binding site by NMR spectroscopy. a, Top, 2D [^{15}N,^{1}H]TROSY spectrum of 0.38 mM CckA-CA recorded at 20 °C. The sequence-specific resonance assignments are indicated. Bottom left, sequence-specific secondary backbone ^{13}C chemical shifts of CckA-CA relative to the random coil values of Kjaergaard et al.[44]. A 1–2–1 smoothing function was applied to the raw data. Consecutive stretches with positive and negative values indicate α-helical and β-strand secondary structure, respectively. The secondary structure elements inferred from these data are indicated above. Asterisks indicate unassigned residues. Bottom right, profile–profile alignment of the CA domains of CckA and DivL carried out with HHpred[45] and formed the basis for the generation of the CckA homology model (shown in Fig. 2G) using the Modeller software[46]. The sequence identity is 25%. Secondary structure elements of CckA as determined by ^{13}Cα secondary chemical shifts and of DivL, as derived from the crystal structure (PDB ID: 4Q20) are shown below the sequence alignment. The residue numbering of CckA is indicated. **b**, Chemical shift perturbation of CckA-CA backbone amide moieties upon c-di-GMP binding. Left, combined chemical shift changes of amide moieties, Δδ(HN), are plotted against the residue number. The magnitudes of 1 s.d. (0.021 p.p.m.) and 2 s.d. (0.042 p.p.m.) are indicated by a purple and pink line, respectively. The arrow points to residues I524 and H525 that experience intermediate chemical exchange upon c-di-GMP binding. Asterisks indicate unassigned residues. Right, region of a 2D [^{15}N,^{1}H]TROSY spectrum of a titration of c-di-GMP to CckA-CA at 20 °C. Sequence-specific resonance assignments are indicated.
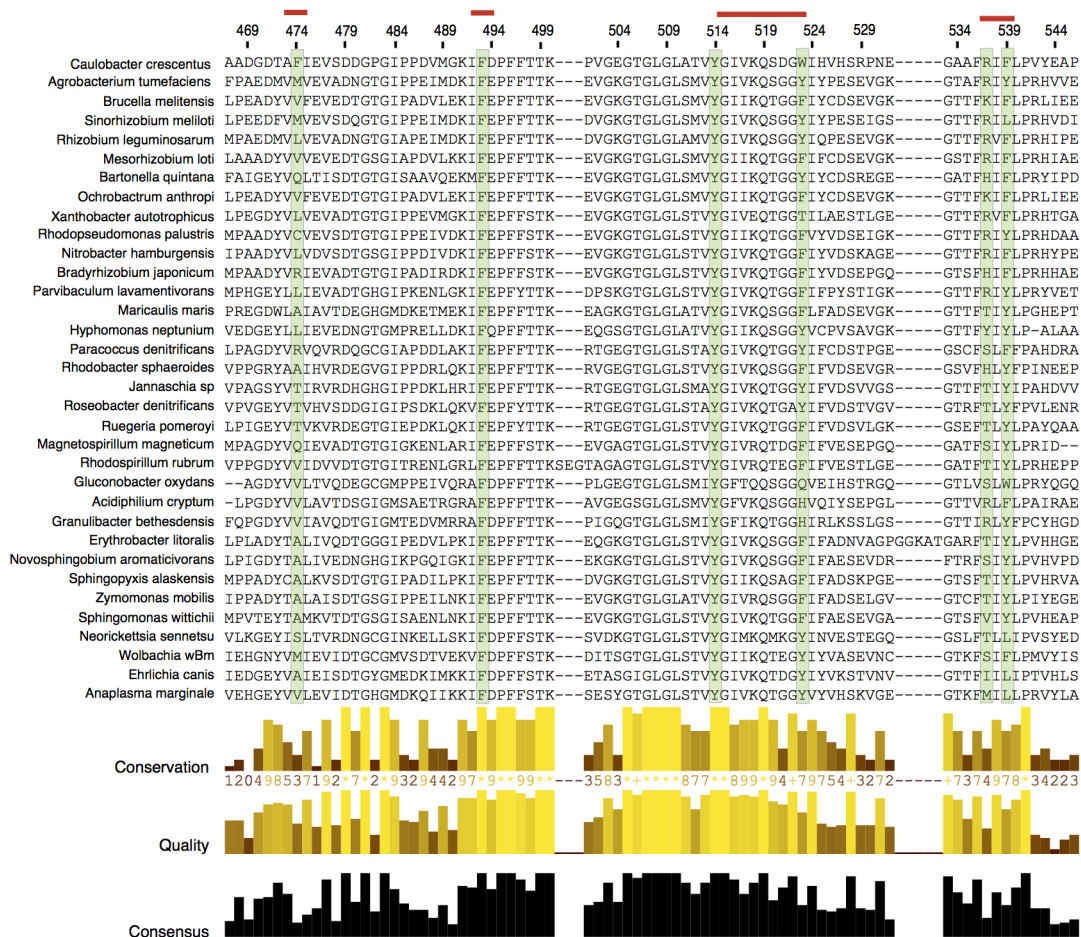
Extended Data Figure 4 | CLUSTALW alignment of the CA domain of CckA. CLUSTALW was used to align the CA domain of CckA from *C. crescentus* and from different alphaproteobacteria. A fragment of the CA domain is shown that corresponds to amino acids 467–546 of *C. crescentus*

CckA. Residues involved in c-di-GMP binding are boxed (green) and red bars above the sequence indicate regions with significant chemical shift in NMR spectroscopy upon c-di-GMP titration. CLUSTALW scores for conservation, quality and consensus are indicated.

a

i) *Wild-type cckA* | ii) *cckA Y514D* | iii) *cckA V366P* | iv) *cckA V366P ΔpleD*

b CtrA occupancy at Cori (fold enrichment)

c

d P*xylX* chromosome

e PYE supplemented with

f Ratio CckA~P/CckA

g Phase / CFP / oriC numbers / cell length unit
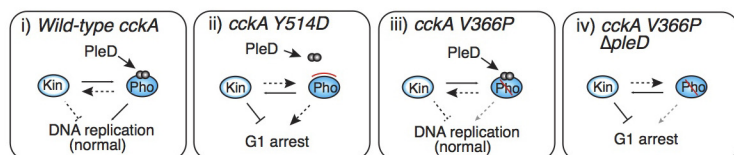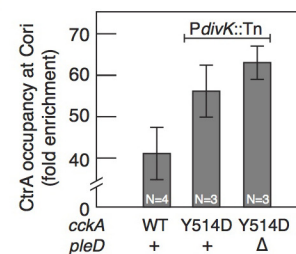
Extended Data Figure 5 | DivK and c-di-GMP convergently control *C. crescentus* growth and replication. a, Model for the regulation of chromosome replication by the CckA kinase/phosphatase switch at reduced levels of DivK. Bold and dotted lines indicate strong and weak reactions, respectively. Dark circles indicate c-di-GMP. The kinase (Kin) and phosphatase modes (Pho) of CckA are indicated. i, C-di-GMP authorizes S-phase entry by inducing CckA phosphatase. ii, C-di-GMP is unable to bind to and induce phosphatase activity of CckA(Y514D) resulting in a G1 arrest. iii, C-di-GMP authorizes S-phase entry by reducing kinase activity of the phosphatase mutant CckA(V366P). iv, Cells lacking PleD fail to downregulate CckA(V366P) kinase activity. **b,** CtrA binding to the origin region is increased in cells containing *cckA(Y514D)* and P*divK*::Tn. CtrA occupancy at the origin was analysed using chromatin immunoprecipitation (ChIP) and quantitative PCR (qPCR) as described in the Methods. Error bars are s.d. **c,** Combining the *cckA(Y514D)* and P*divK*::Tn alleles leads to a G1 arrest. Exponential cultures of mutants containing different combinations of *cckA*, P*divK* and *pleD* alleles were analysed by light microscopy and flow cytometry. Representative examples of two biological replicates for phase-contrast images and profiles of DNA content are shown with scale bars of 5 μm. **d,** Reduced *divK* expression in P*xylX*::*divK* strains containing the *cckA(Y514D)* allele leads to G1 arrest. Top, schematic of the chromosomal arrangement of cells expressing *divK* from the P*xyl* promoter (P*xylX*::*divK*) and harbouring different *cckA* alleles. The *divK* gene is fused to the P*xylX* promoter in the *xylX* locus. The chromosomal copy of *divK* at the original locus was replaced with a Ω cassette. Different *cckA* alleles were introduced at the *cckA* locus by allelic exchange. Bottom left, cellular levels of DivK and PleD as determined by immunoblot analysis in strains grown in the presence or absence of xylose. Cells expressing *divK* from its own promoter at the native locus (P*divK*) were used as control. Note that for P*xylX*::*divK* derivatives grown in peptone yeast extract, twice as many cells were used. Band intensities were determined with ImageJ and the respective values shown as relative units compared to wild type. Bottom right, DNA content per cell mass (DNA/mass) was analysed as described in Fig. 1c and values are shown below the graphs. Fractions of cells containing more than two chromosomes are indicated by brackets. Representatives of two biological replicates are shown. **e,** Colony-forming ability of P*xylX*::*divK* strains carrying different *cckA* alleles. Fivefold serial dilutions of the indicated strains were spotted and grown for 2 days at 30 °C on peptone yeast extract plates with the supplements indicated. Representatives of two biological replicates are shown. Note tha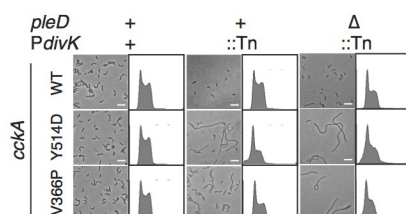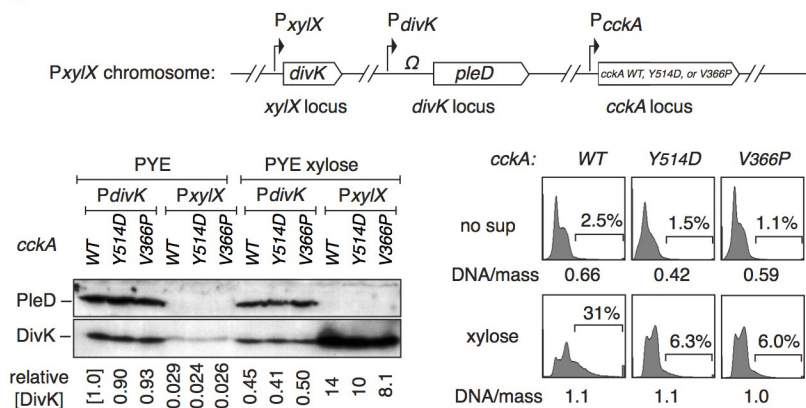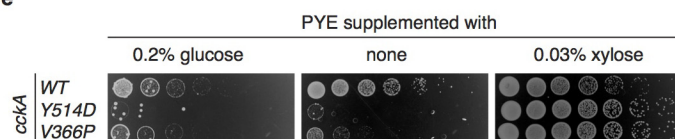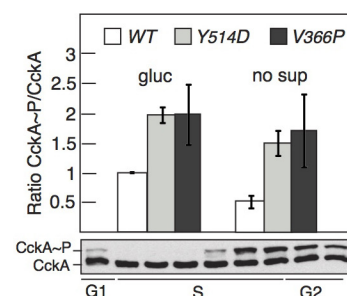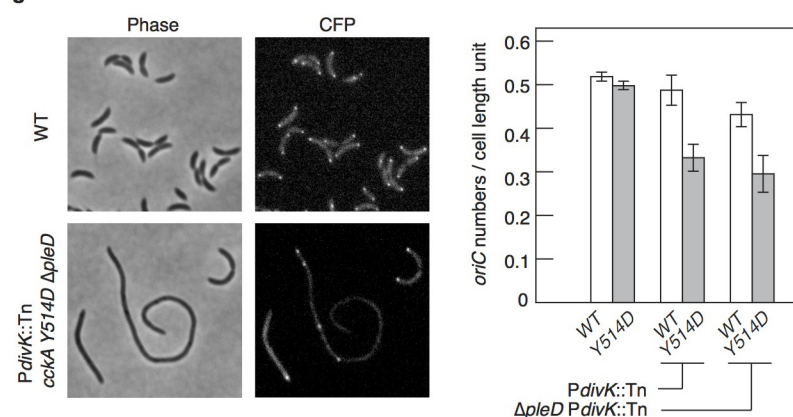t these results are consistent with individual DNA replication profiles shown in Fig. 3 and panel **d**. **f,** Reduced *divK* expression in P*xylX*::*divK* strains containing the *cckA(Y514D)* allele leads to increased CckA phosphorylation levels. The cellular fraction of phosphorylated CckA (CckA~P) was determined using Phos-tag gel electrophoresis. As a control, CckA and phosphorylated CckA levels were determined in synchronized populations of wild-type cells proceeding through the cell cycle (bottom). P*xylX*::*divK* strains harbouring different *cckA* alleles were analysed during exponential growth at 30 °C in the presence or absence of glucose (0.3%). The addition of glucose reduces leaky expression from the P*xyl* promoter. Relative ratios of phosphorylated CckA to total CckA are shown. Error bars are s.d. Averages and standard deviations were obtained from three biological replicates. **g,** Single-cell analysis of the replication status of mutants with reduced DivK levels and abolished CckA control by c-di-GMP. Strains producing LacI–CFP and harbouring an array of *lac* operator (*lacO*) sites near the origin of replication were analysed[32]. Fluorescent repressor–operator system strains contained wild-type *cckA* or the *cckA(Y514D)* mutant allele, as well as P*divK*::Tn-tet with wild-type *pleD* or Δ*pleD* as indicated. Representative phase-contrast and fluorescence images of two biological replicates are shown. Numbers of origins per cell length units were analysed statistically and the mean value and standard deviations obtained from two biological replicates are shown as a column graph. For each strain, a total of >900 cells were analysed using MicrobeTracker (http://microbetracker.org). Note that these results are consistent with the DNA replication profiles of equivalent strains without the FROS module shown in Fig. 3 and panels **c** and **d**.
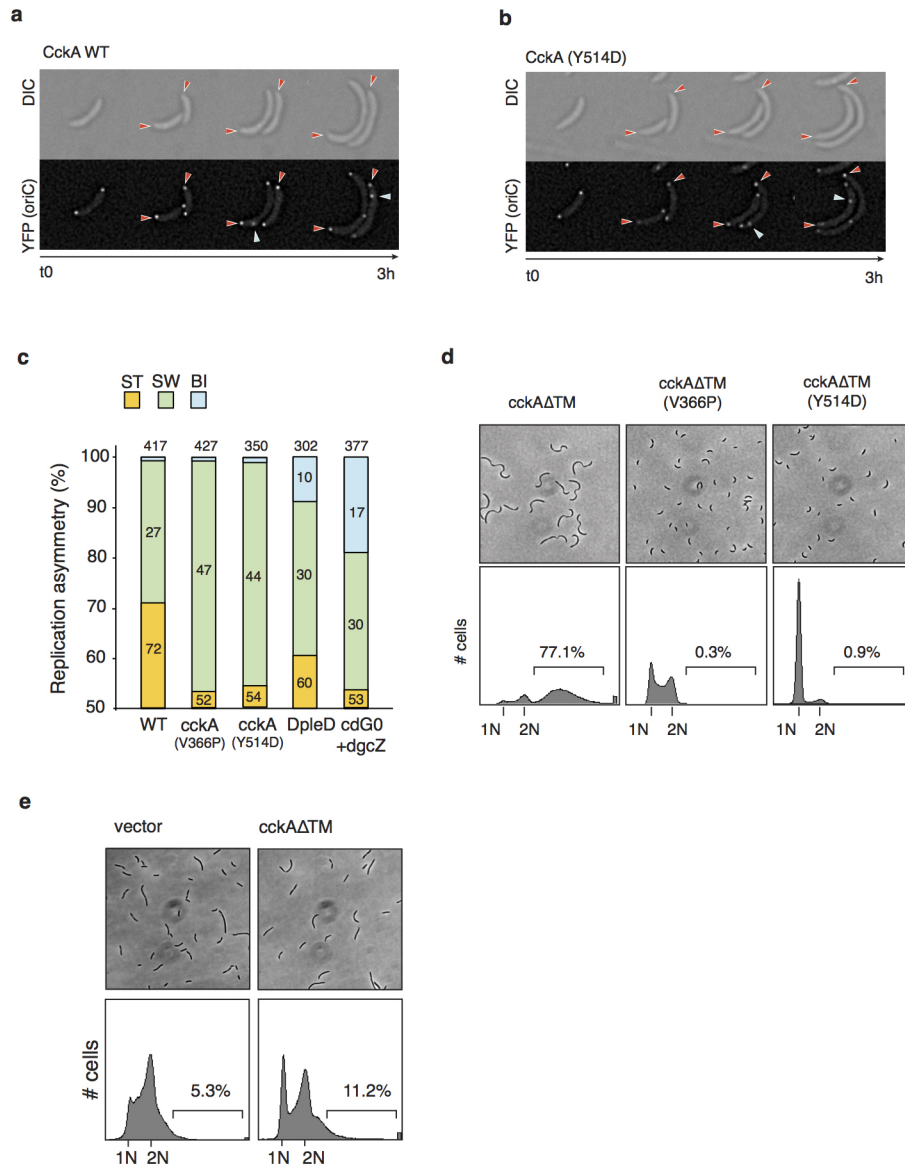
Extended Data Figure 6 | C-di-GMP-mediated spatial control of CckA directs replication asymmetry in dividing cells. a, b, Fluorescent repressor–operator system analysis was used to visualize DNA replication in individual cells. Dividing cells of wild-type *C. crescentus* (**a**) and *cckA(Y514D)* mutant (**b**) producing TetR–YFP and harbouring an array of tet operator (*tetO*) sites near the origin of replication were analysed by fluorescence microscopy. Frames from representative time-lapse movies used for panel **c** are shown. Stalked/old poles of newly divided daughter cells are marked with red arrows; newly replicated origins are marked with blue arrows. **c,** Spatial patterns of DNA replication were scored using a Tet-based fluorescent repressor–operator system and divided into three classes as indicated: replication initiation at the origin of replication located at the stalked pole (ST, orange), swarmer pole (SW, green), or at both poles (BI, blue). The bar diagram shows the quantification of

wild-type and mutant strains with numbers indicating the percentage of cells falling into the three classes. The total number of cells analysed (*n*) is indicated above each bar. **d,** Expression *cckAΔTM* leads to c-di-GMP-dependent over-replication. Wild-type *C. crescentus* strains expressing different *cckAΔTM* alleles were analysed by light microscopy and flow cytometry as indicated. The fraction of cells bearing more than two chromosomes is indicated and shown as percentage. Representatives of two biological replicates are shown. **e,** Expression *cckAΔTM* leads to c-di-GMP-dependent over-replication. A *C. crescentus* cdG0 strain expressing *cckAΔTM* was analysed by light microscopy and flow cytometry as indicated. The fraction of cells bearing more than two chromosomes is indicated and shown as percentage. Representatives of two biological replicates are shown.

LETTER

doi:10.1038/nature14450

Condensin–driven remodelling of X chromosome topology during dosage compensation

Emily Crane[1]†*, Qian Bian[1]*, Rachel Patton McCord[2]*, Bryan R. Lajoie[2]*, Bayly S. Wheeler[1], Edward J. Ralston[1], Satoru Uzawa[1], Job Dekker[2] & Barbara J. Meyer[1]

The three-dimensional organization of a genome plays a critical role in regulating gene expression, yet little is known about the machinery and mechanisms that determine higher-order chromosome structure[1,2]. Here we perform genome-wide chromosome conformation capture analysis, fluorescent *in situ* hybridization (FISH), and RNA-seq to obtain comprehensive three-dimensional (3D) maps of the *Caenorhabditis elegans* genome and to dissect X chromosome dosage compensation, which balances gene expression between XX hermaphrodites and XO males. The dosage compensation complex (DCC), a condensin complex, binds to both hermaphrodite X chromosomes via sequence-specific recruitment elements on X (*rex* sites) to reduce chromosome-wide gene expression by half[3–7]. Most DCC condensin subunits also act in other condensin complexes to control the compaction and resolution of all mitotic and meiotic chromosomes[5,6]. By comparing chromosome structure in wild-type and DCC-defective embryos, we show that the DCC remodels hermaphrodite X chromosomes into a sex-specific spatial conformation distinct from autosomes. Dosage-compensated X chromosomes consist of self-interacting domains (~1 Mb) resembling mammalian topologically associating domains (TADs)[8,9]. TADs on X chromosomes have stronger boundaries and more regular spacing than on autosomes. Many TAD boundaries on X chromosomes coincide with the highest-affinity *rex* sites and become diminished or lost in DCC-defective mutants, thereby converting the topology of X to a conformation resembling autosomes. *rex* sites engage in DCC-dependent long-range interactions, with the most frequent interactions occurring between *rex* sites at DCC-dependent TAD boundaries. These results imply that the DCC reshapes the topology of X chromosomes by forming new TAD boundaries and reinforcing weak boundaries through interactions between its highest-affinity binding sites. As this model predicts, deletion of an endogenous *rex* site at a DCC-dependent TAD boundary using CRISPR/Cas9 greatly diminished the boundary. Thus, the DCC imposes a distinct higher-order structure onto X chromosomes while regulating gene expression chromosome-wide.

To compare the molecular topology of X chromosomes and autosomes in *C. elegans*, we generated genome-wide chromatin interaction maps from mixed-stage embryos using a modified chromosome conformation capture (Hi-C) protocol combining conventional chromosome conformation capture (3C) with paired-end sequencing[10–12] (Fig. 1, Extended Data Fig. 1 and Methods). Interaction data, binned at both 10 kb and 50 kb intervals, revealed features observed in other organisms. Interactions occur most frequently in *cis* and decay with genomic distance (Extended Data Fig. 1 and Methods). Chromosome compartments comparable to active A and inactive B compartments[11,13] are formed (Extended Data Figs 1 and 4–6). Compartments at the left end of the X chromosome and both ends of autosomes align with binding domains for lamin[14], lamin-associated protein LEM-2 (Extended Data
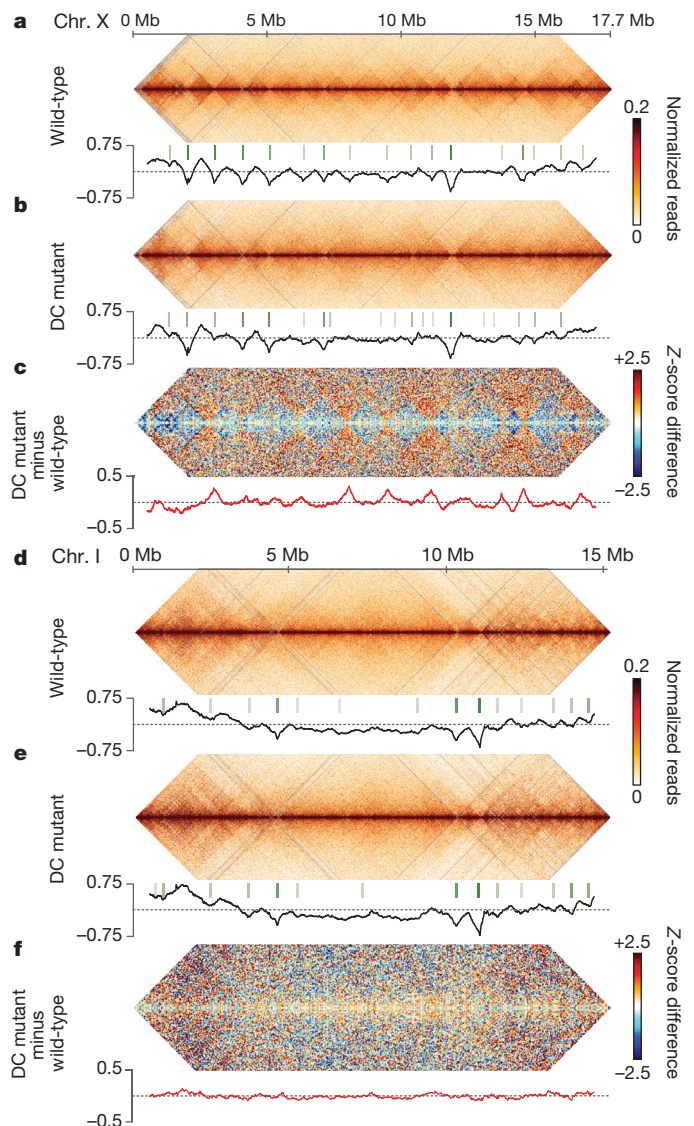
Figure 1 | DCC modulates spatial organization of X chromosomes.
a, **b**, **d**, **e**, Chromatin interaction maps binned at 10 kb resolution show interactions 0–4 Mb apart on chromosomes X and I in wild-type and DC mutant embryos. Plots (black) show insulation profiles. Minima (green lines) reflect TAD boundaries. Darker green indicates stronger boundary. **c**, **f**, Blue–red *Z*-score difference maps binned at 50 kb resolution for X and I show increased (orange–red) and decreased (blue) chromatin interactions between mutant and wild-type embryos. Differential insulation plots (red) show insulation changes between mutant and wild-type embryos.

[1]Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California-Berkeley, Berkeley, California 94720-3204, USA. [2]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA. †Present address: Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5120, USA.
*These authors contributed equally to this work.
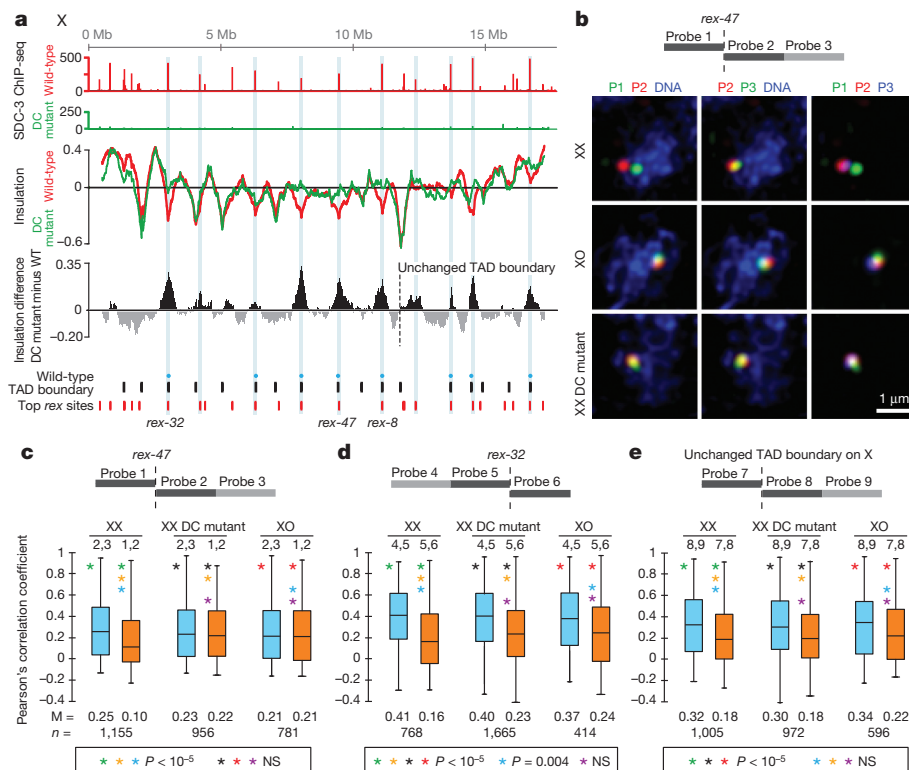
Figure 2 | FISH shows DCC-dependent TAD boundaries at high-affinity *rex* sites. a, High DCC occupancy correlates with TAD boundaries lost or reduced upon DCC depletion. Top, ChIP-seq profiles of DCC subunit SDC-3 in wild-type (red) and DC mutant (green) embryos. The *y* axis, reads per million (RPM) normalized to IgG control. Middle, insulation profiles of wild-type (red) and DC mutant (green) embryos. Bottom, insulation difference plot for wild-type insulation profile subtracted from DC mutant profile. Black lines, TAD boundary locations. Blue dots, boundaries with insulation changes >0.1 between wild-type and DC mutant embryos. Red lines, locations of 25 highest DCC-occupied *rex* sites. Cyan bars, sites with the largest insulation loss. **b**, Confocal images of embryonic nuclei of various genotypes stained with a DNA intercalating dye (blue) and 500 kb FISH probes around the *rex-47* TAD boundary. **c, d, e**, Quantification of FISH probe colocalization confirms DCC-dependent and DCC-independent boundaries found by Hi-C. Box plots, distribution of Pearson's correlation coefficients between pairwise combinations of FISH probes within (blue) or across (orange) TADs. Boxes, middle 50% of coefficients. Centre bars, median (M) coefficients. *n*, total number of nuclei. Asterisks of same colour specify data compared using one-tailed Mann–Whitney *U*-test. NS, not significant.

Figs 4–6)[15], and the H3K9me3 inactive chromatin mark[16], suggesting their similarity to inactive B compartments of mammals.

Chromatin interaction maps also revealed self-interacting domains (∼1 Mb), predominantly on X chromosomes. These domains are visible as diamonds along the interaction maps (Fig. 1a, d) and resemble TADs of mammalian and fly chromosomes[8,9,12]. To quantify TADs, we devised an approach of assigning an 'insulation score' to genomic intervals along the chromosome. The score reflects the aggregate of interactions occurring across each interval. Minima of the insulation profile denote areas of high insulation we classified as TAD boundaries (Methods, Fig. 1, Extended Data Figs 2a and 3a, b).

The insulation profile of the X chromosome stands out compared to those of autosomes. The insulation signal amplitude is larger on the X chromosome (Fig. 1a, d and Extended Data Fig. 3d), implying TAD boundaries are stronger. Also, TAD boundaries on the X chromosome are more abundant and regularly spaced (Extended Data Fig. 3d).

To assess whether the DCC controls the spatial organization of herm-aphrodite X chromosomes, we generated chromatin interaction maps for a dosage-compensation-defective mutant (DC mutant; Fig. 1 and Extended Data Figs 1–6) in which the XX-specific DCC recruitment factor SDC-2 was depleted, severely reducing DCC binding to X[3,4,17] (Fig. 2a) and elevating X chromosome gene expression (see below). The insulation profile of the X chromosome, but not autosomes, was greatly changed (Fig. 1b, e and Extended Data Figs 1–6). Of a total of 17 TAD boundaries on the X chromosome, 5 were eliminated and 3 severely reduced in insulation. TAD boundary strength and spacing on the X chromosome in DC mutants resembled that of autosomes (Extended Data Fig. 3d).

To characterize this transformation in conformation, we calculated the difference between chromatin interaction maps of wild-type and DC mutant embryos after converting the interaction data into genomic-distance-normalized *Z*-scores. In DC mutants, interactions on X increased across TAD boundaries but decreased within TADs, revealing a DCC-dependent remodelling of X chromosome structure (Fig. 1c and Extended Data Figs 1–3 and 5). Weakening of TAD boundaries is expected to cause chromosome-wide changes in chromatin interactions. The largest changes in insulation on X occurred at TAD boundaries. Autosomes appeared unaffected (Figs 1c, f and 2a and Extended Data Figs 1–4 and 6).

TAD boundaries on the X chromosome are enriched for the highest DCC-occupied *rex* sites[3,4,18] (Fig. 2a and Extended Data Fig. 7d). About 50% of all TAD boundaries and 90% of changed ones overlap the top 25 *rex* sites, a correlation higher than expected at random (Extended Data Fig. 7d). In DC mutants, the largest insulation losses occurred in regions overlapping the strongest *rex* sites (Fig. 2a). These results imply the DCC plays a direct role in defining TADs by binding to *rex* sites to mediate formation of TAD boundaries. In contrast, genomic features such as highly occupied targets (HOT) sites[19] do not govern TADs (Supplementary Table 2).

Two TAD boundaries on X that overlap *rex* sites in the LEM-2 B-like compartment were not greatly reduced in DC mutants (Figs 1 and 2a and Extended Data Fig. 5e). Although the DCC exerts a dominant influence on TAD formation, other forces act on the X chromosome to form TADs, as on autosomes.

To confirm the DCC-dependent topology of the X chromosome, we visualized TADs using quantitative 3D fluorescent *in situ* hybridization (FISH) in wild-type XX embryos and embryos lacking DCC binding on X: male XO and DC-mutant XX (Fig. 2b–e). We imaged fluorescent probes that tiled 500 kb regions within TADs or flanking TAD boundaries. Probe overlap was quantified by analysing the distribution of Pearson's correlation coefficients between FISH signals from pairwise probe combinations[8].

As expected for TADs in wild-type embryos, two adjacent probes within a TAD on either X chromosomes or autosomes overlapped to a greater extent than two adjacent probes on either side of a TAD boundary (Fig. 2b–e and Extended Data Fig. 8a–d). For DCC-dependent TAD boundaries on X including *rex-47*, *rex-32* and *rex-8*, adjacent probes flanking TAD boundaries overlapped and colocalized more in embryos lacking DCC binding than in wild-type XX embryos (Fig. 2c, d and Extended Data Fig. 8b). In contrast, the DCC-independent TAD boundaries on the X chromosome and autosomes did not change (Fig. 2e and Extended Data Fig. 8c, d). FISH analysis also confirmed that some DCC-dependent TAD boundaries were eliminated (*rex-47*), and others reduced (*rex-32*) in DC mutants and XO males (Fig. 2c, d), showing that the DCC alters X chromosome structure by strengthening pre-existing TAD boundaries and creating new ones.

Figure 3 | Strong DCC-dependent interactions occur between high-affinity *rex* sites at TAD boundaries. a, Cumulative distribution of Hi-C Z-scores for interactions between 10 kb bins with *rex* sites or with other X chromosome interactions in wild-type or DC mutant embryos. Interactions >4 Mb were excluded from panels **a–e**. P values are corrected for multiple testing. In wild-type embryos, *rex–rex* interactions are stronger than all other X chromosome interactions ($P < 2 \times 10^{-16}$; two-sided KS test) and stronger than *rex–rex* interactions in DC mutants ($P = 1.5 \times 10^{-9}$; Wilcoxon signed rank test). **b**, Distributions of Hi-C Z-scores show that *rex–rex* interactions are stronger than non-*rex* interactions ($P < 2 \times 10^{-16}$; two-sided KS test) or *rex* to non-*rex* interactions ($P = 1.7 \times 10^{-14}$; two-sided KS test). **c**, Distributions of Z-score differences (DC mutant minus wild-type) show that *rex–rex* interactions decrease more than any of 1,000 random sets of non-*rex* interactions of equal number ($P < 0.001$). **d**, Average Hi-C interaction profiles (normalized read counts) around pairs of top 25 *rex* sites or all known *rex* sites, in wild-type and DC mutant embryos. *rex* sites are centred at 0. **e**, Distributions of Hi-C Z-scores for interactions between bins with *rex* or non-*rex* sites at TAD boundaries or within TADs of wild-type (left) or DC mutant (middle) embryos. *rex* sites interact more at TAD boundaries than in TADs ($P = 0.0025$). These sets of interactions are not different in DC mutants ($P = 0.348$). Interactions at TAD boundaries or within TADs on autosomes (right). **f**, Circos plots depict all *rex–rex* interactions (Z-score >2, coloured line) in 50 kb bins in wild-type embryos. Concentric circles show insulation difference plot (black and grey), wild-type TAD boundaries (green boxes), and *rex* sites (black lines, strongest sites named). **g**, *rex–rex* interactions in **f** that are retained in DC mutants. **h**, Deletion of *rex-47* disrupts TAD boundary. Box plots of Pearson's correlation coefficients for FISH probe combinations in wild-type, *rex-47* Δ, and DC mutant. Probe overlap across TAD boundary increased in *rex-47* Δ vs. wild-type ($P < 0.01$ ANOVA) but was not different in *rex-47* Δ vs. DC mutants ($P = $ NS, ANOVA). Probe overlap in TAD was not different in 3 strains ($P = 0.075$, ANOVA).

Robust correlation between *rex* sites, DCC-dependent TAD boundaries, and regions of greatest insulation loss in DC mutants (Fig. 2a, Extended Data Fig. 7d and Supplementary Table 2) led us to test whether *rex* sites interact in a DCC-dependent manner. We found *rex–rex* interactions to be among the most prominent interactions on the X chromosome by comparing the ranking (Extended Data Fig. 7a) and cumulative distribution (Fig. 3a, b) of Z-scores for *rex* interactions with those for all other X chromosome interactions. In DC mutants, *rex–rex* interactions decreased more than any of the 1,000 random sets of X chromosome interactions (Fig. 3a, c and Extended Data Fig. 7b, c, e). These observations support the hypothesis that DCC binding at *rex* sites facilitates *rex–rex* interactions.

The *rex–rex* interaction frequency was directly related to the level of DCC occupancy at *rex* sites, as shown by 3D profiles of Hi-C interaction frequencies made for pairwise combinations of 10 kb bins overlapping either the top 25 DCC-occupied *rex* sites or all 64 *rex* sites (Fig. 2a and 3d, Extended Data Fig. 7f and Supplementary Table 2). Interactions for the top 25 *rex* sites exceeded those for all *rex* sites.

The correlation between *rex*-interaction strength and DCC occupancy was reinforced by contrasting results with dependent on X (*dox*) sites. The DCC spreads to these lower affinity *dox* sites located in promoters of highly expressed genes once recruited to X by *rex* sites[3,4]. *dox* sites showed no substantial interactions in 3D plots (Extended Data Fig. 7g).

The strongest *rex–rex* interactions occurred between *rex* sites at DCC-dependent TAD boundaries on the X chromosome (Fig. 3e). Weaker *rex–rex* interactions also occur within TADs. In DC mutants, *rex* interactions within TADs and between TAD boundaries diminished to the level of non-*rex* interactions (Fig. 3e). For autosomes, in contrast, interactions between TAD boundaries were not greater than interactions within TADs, and neither set of interactions changed in DC mutants (Fig. 3e and Extended Data Fig. 7h). These results suggest that DCC-dependent interactions between *rex* sites at TAD boundaries contribute more to boundary formation on X than *rex* interactions within TADs, although DCC-dependent *rex* interactions within TADs might contribute to TAD integrity.

Visualization of Hi-C interaction data via Circos plots shows that almost all *rex* sites engage in one or multiple strong DCC-dependent interactions with other *rex* sites, particularly at adjacent TAD boundaries (Fig. 3f, g). Together, our findings reinforce the model that *rex* sites contribute to TAD formation by recruiting the DCC and facilitating DCC-dependent looping interactions between *rex* sites at TAD boundaries. In contrast, TAD boundaries on autosomes do not appear to result from looping interactions between boundaries (Fig. 3e, right panel and Extended Data Fig. 7h), suggesting that different strategies govern, in part, the formation of DCC-dependent and autosomal TADs.
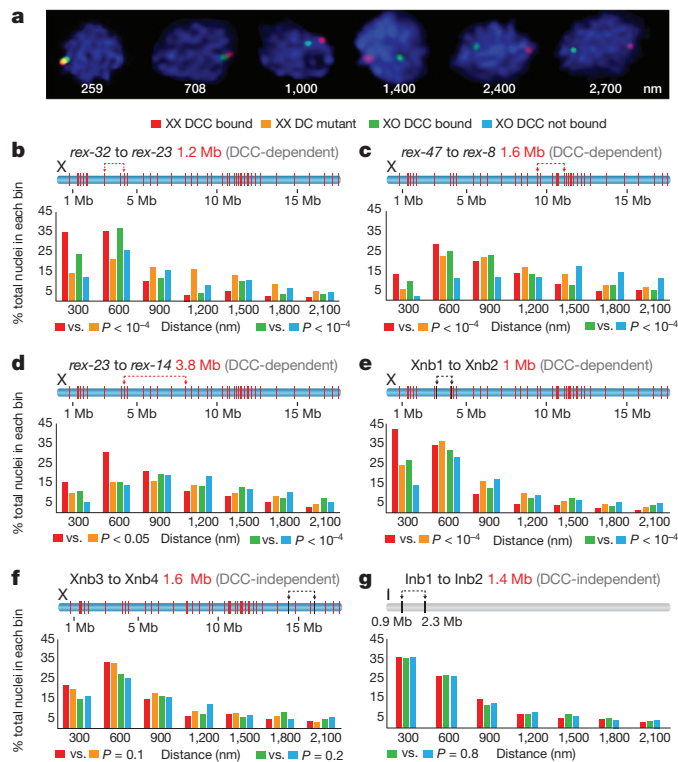
Figure 4 | Quantitative FISH shows DCC-dependent association of *rex* sites in single cells. a, Representative embryonic nuclei show variability in spacing of FISH probes (red, green) targeting two *rex* sites. **b–g**, Quantification of the 3D distance between FISH probes in embryos of different genotypes. DCC binding to the single X chromosome of XO embryos was achieved using an XO lethal (*xol-1*) mutation, which activates *sdc-2*, the XX-specific trigger of DCC assembly[20]. Total number of nuclei is given in Extended Data Fig. 9a–f. **b–d**, Pairs of *rex* sites at DCC-dependent TAD boundaries of varying genomic separation. **e**, A pair of sites on the X chromosome that lack DCC binding sites within 100 kb but have DCC-dependent Hi-C interactions. **f, g**, Loci on chromosome X and chromosome I that lack DCC binding sites within 80–90 kb and display DCC-independent Hi-C interactions. **b–g**, Distances between FISH spots were binned in 300 nm intervals and represented in relative frequency histograms. Schematic above each histogram depicts the locations of FISH probes (arrows), their genomic separation (red text), and the location of all *rex* sites (red bars) or sites lacking DCC binding (black). The DCC dependence or independence of the corresponding Hi-C interactions is indicated above the histogram (grey). *P* values comparing genotypes were calculated using the chi-square test to compare the 0–300 nm bin with 301–2,700 nm bins. The 0–300 nm bin contains FISH probes considered co-localized, because probes <300 nm apart always overlap visually, while probes 700 nm apart appear only adjacent to each other.

The model that *rex* interactions play a critical role in establishing and reinforcing TAD boundaries makes specific predictions. First, *rex* interactions identified by Hi-C should be evident by FISH. Second, deletion of a strong *rex* site from a DCC-dependent TAD boundary should reduce or eliminate the boundary. Both predictions were verified by the data.

To confirm DCC-dependent *rex–rex* interactions and further assess X-chromosome topology, we devised a FISH assay using 3–6 kb probes to quantify the spatial separation between two sites (Methods and Fig. 4). We compared distances between loci in XX embryos with (wild-type) and without (DC mutant) DCC binding on the X chromosome to quantify the level and DCC-dependence of interactions. We also compared distances in XO embryos with and without DCC binding on the X chromosome to quantify DCC-dependent interactions that occur between loci on the same chromosome (Fig. 4 legend). Hi-C analysis did not distinguish between interactions within the same chromosome or across homologous chromosomes.

FISH analysis confirmed all categories of interactions shown by Hi-C: (1) strong DCC-dependent interactions between *rex* sites at DCC-dependent TAD boundaries (*rex-32* to *rex-23*, *rex-47* to *rex-8*, and *rex-23* to *rex-14*); (2) strong DCC-dependent interactions between X loci lacking DCC binding (Xnb1 to Xnb2 and Xnb7 to Xnb8 (nb, not bound)); (3) strong DCC-independent interactions between loci on X (Xnb3 to Xnb4) or I (Inb1 to Inb2) that lacked DCC binding; and (4) weak DCC-independent interactions between distant loci on X (Xnb5 to Xnb6) or I (Inb3 to Inb4) that lack DCC binding (Fig. 4b–g and Extended Data Fig. 9a–f, i–k). FISH and Hi-C results agreed, for both the strength and DCC-dependence of interactions (Extended Data Fig. 9g, h).

The only discrepancy occurred for distantly spaced *rex* loci (*rex-1* to *rex-8* (6.7 Mb); *rex-32* to *rex-8* (8.1 Mb)), which showed greater DCC-dependent spatial proximity by FISH analysis than predicted by Hi-C (Extended Data Fig. 9l, m). Loss of sensitivity in our Hi-C data for sites separated by >5 Mb may account for the difference.

Both FISH and Hi-C experiments showed that the DCC-dependent topology of the X chromosome brings many distant, non-*rex* sites into close proximity. If the DCC compacted the X chromosome uniformly, pairs of non-*rex* loci separated by similar distances should exhibit comparable levels of DCC-dependent interactions. However, they did not. For example, two pairs of non-*rex* loci (Xnb1 and Xnb2 (1 Mb); Xnb7 and Xnb8 (1.4 Mb)) showed strong DCC-dependent interactions (Fig. 4e and Extended Data Fig. 9g, h, k), but the non-*rex* loci Xnb3 and Xnb4 (1.6 Mb) showed strong DCC-independent interactions (Fig. 4f). Thus, the DCC affects the overall topology of the X chromosome but does not cause uniform compaction across the X chromosome.

To test whether DCC-dependent interactions between *rex* sites create TAD boundaries, we deleted the endogenous *rex-47* site from a DCC-dependent TAD boundary using genome editing with CRISPR/Cas9 (Extended Data Fig. 8e, f) and assayed TAD structure with FISH (Fig. 3h). Chromatin immunoprecipitation followed by quantitative polymerase chain reaction (ChIP–qPCR) showed the deleted *rex* locus (*rex-47* Δ) lacked DCC binding (Extended Data Fig. 8g). The TAD boundary was greatly diminished, as predicted (Fig. 3h). For FISH probes flanking the *rex-47* TAD boundary, overlap was increased in *rex-47* Δ and DC mutant embryos over that in wild-type embryos. In contrast, overlap was not statistically different between *rex-47* Δ and DC mutant embryos. Thus, the DCC plays a key role in inducing and reinforcing TAD boundaries on X by mediating long-range interactions between its highest-affinity *rex* sites.

We explored the relationship between TAD structure and gene expression. Our prior work showed the DCC acts at a distance to repress gene expression[3,4,20], suggesting that a unique, DCC-dependent X-chromosome structure might mediate chromosome-wide gene repression, as supported by our Hi-C and FISH data. We assessed whether the structure of individual TADs affects gene expression locally or whether the chromosome-wide topology created from TADs regulates gene expression globally. Both RNA-seq data derived from embryo preparations used for Hi-C analysis and GRO-seq data from independent embryo preparations support the latter hypothesis for the following reasons.

First, in wild-type embryos, genes at TAD boundaries were not expressed at significantly different levels from genes within TADs, for either chromosome X (Fig. 5b and Extended Data Fig. 10a, d) or chromosome I (Fig. 5f). Second, although the X chromosome is organized into DCC-dependent TADs in wild-type animals, no similarly coordinated block of genes exhibited elevated expression in DC mutants (Fig. 5a). That is, the changes in expression were not significantly different for X-linked genes within TADs, at all TAD boundaries, at changed TAD boundaries, or within regions of changed insulation (Fig. 5c, d and Extended Data Fig. 10b, c, e–i). Similarly, DC mutations did not alter gene expression on chromosome I in any discernible pattern (Fig. 5e, g, h and Extended Data Fig. 10g–i).

Our results support the model that TAD structure on the X chromosome mediated by DCC binding to *rex* sites creates a 3D topology that acts chromosome-wide to repress gene expression. Given that changes in TAD boundaries occur locally, while changes in gene expression occur chromosome-wide, a parsimonious model posits that
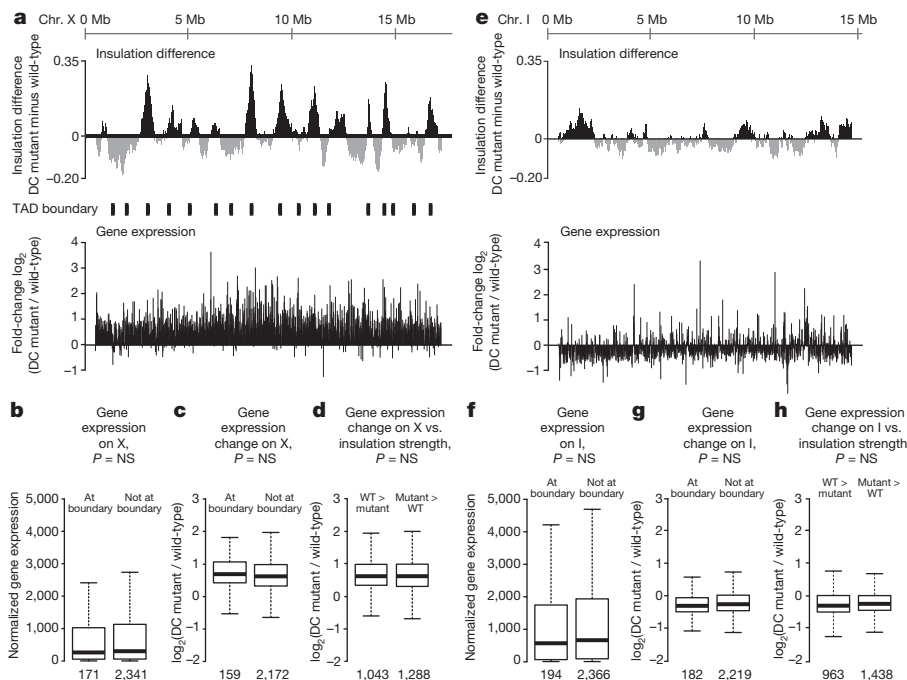
Figure 5 | DCC-dependent TADs influence global rather than local gene expression.
a, Insulation changes and TAD boundaries are compared to median fold-changes in expression (10 kb bins across chromosome X) between wild-type and DC mutant embryos. **a–h**, No discernible pattern was detected between mutant-induced changes in expression and gene locations relative to TADs or TAD boundaries. **b**, Box plots show comparison of expression levels for X chromosome genes within or outside TAD boundaries in wild-type embryos. Expression levels, normalized read number per kilobase of gene length. **c**, Box plots show expression changes for X chromosome genes within or outside TAD boundaries. **d**, Comparison of expression changes (DC mutant/wild-type) for X chromosome gene sets with greater insulation scores in wild-type embryos (grey domains in **a**) versus in DC mutants (black domains in **a**). **e–h**, Same as **a–d** but for genes on chromosome I. P values for **b–d** and **f–h**, Mann–Whitney U-test; no significant P values withstood multiple testing correction. NS, not significant. **a**, **c**, **d**, **e**, **g**, **h**, Lowest-expressed genes (bottom 10%) were removed from analyses.

DCC-dependent changes in X chromosome structure imposed by *rex–rex* interactions drive the chromosome-wide reduction in gene expression. Potential DCC-dependent nuclear positioning of the X chromosome might also affect gene expression, as speculated by others[21].

In summary, DCC-induced formation of TAD structure on the X chromosome demonstrates a striking remodelling of chromosome topology that reveals a central role for condensin in shaping the 3D landscape of interphase chromosomes. Not only does condensin compact and resolve mitotic and meiotic chromosomes, it acts as a key structural element to regulate gene expression. No other molecular complex or set of DNA binding sites is yet known to cause comparably strong effects on megabase-scale TAD structure in higher eukaryotes[22–24]. Our new understanding of the topology of dosage-compensated chromosomes provides fertile ground to decipher the detailed mechanistic relationship between higher-order chromosome structure and chromosome-wide regulation of gene expression.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 July 2014; accepted 7 April 2015.
Published online 1 June 2015.

1. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
2. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
3. Jans, J. *et al.* A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes Dev.* **23**, 602–618 (2009).
4. Pferdehirt, R. R., Kruesi, W. S. & Meyer, B. J. An MLL/COMPASS subunit functions in the *C. elegans* dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes Dev.* **25**, 499–515 (2011).
5. Csankovszki, G. *et al.* Three distinct condensin complexes control *C. elegans* chromosome dynamics. *Curr. Biol.* **19**, 9–19 (2009).
6. Mets, D. G. & Meyer, B. J. Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. *Cell* **139**, 73–86 (2009).
7. Meyer, B. J. Targeting X chromosomes for repression. *Curr. Opin. Genet. Dev.* **20**, 179–189 (2010).
8. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
9. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
10. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
11. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
12. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
13. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).
14. Towbin, B. D. *et al.* Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934–947 (2012).
15. Ikegami, K., Egelhofer, T. A., Strome, S. & Lieb, J. D. *Caenorhabditis elegans* chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol.* **11**, R120 (2010).
16. Liu, T. *et al.* Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.* **21**, 227–236 (2011).
17. Dawes, H. E. *et al.* Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate. *Science* **284**, 1800–1804 (1999).
18. McDonel, P., Jans, J., Peterson, B. K. & Meyer, B. J. Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**, 614–618 (2006).
19. Chen, R. A. *et al.* Extreme HOT regions are CpG-dense promoters in *C. elegans* and humans. *Genome Res.* **24**, 1138–1146 (2014).
20. Kruesi, W. S., Core, L. J., Waters, C. T., Lis, J. T. & Meyer, B. J. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *Elife* **2**, e00808 (2013).
21. Sharma, R. *et al.* Differential spatial and structural organization of the X chromosome underlies dosage compensation in *C. elegans*. *Genes Dev.* **28**, 2591–2596 (2014).
22. Seitan, V. C. *et al.* Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* **23**, 2066–2077 (2013).
23. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
24. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).

Author Contributions E.C. conducted Hi-C, ChIP-seq and FISH experiments. Q.B. conducted FISH and *rex-47* deletion experiments. R.P.M. conducted statistical and long-range interaction analyses. B.R.L. analysed Hi-C data and mapped TADs. B.S.W. conducted RNA-seq studies. E.J.R., S.U. and all authors analysed data and edited the manuscript, J.D. guided and performed Hi-C analysis and wrote manuscript sections. B.J.M. guided the study and wrote the manuscript.

METHODS

Nematode strains. The strains used in this study are as follows. Wild-type: TY125, N2 Bristol, XX. Dosage compensation mutants: *sdc-2 (y93, RNAi)* X (XX strain used in all experiments requiring a DC mutant strain, except those listed below using TY2222 or TY1996); TY1996, *szT1/sdc-2(y74) unc-3(e151)* X (XX DC mutant in Figs 2b–e, 3h and 4b–f and Extended Data Fig. 9a–e, i); TY2222, *her-1(hv1y101) V; xol-1(y9) sdc-2(y74) unc-9(e101)* X (XX DC mutant used only in Extended Data Fig. 9j); TY0810, *sdc-2(y93)* X (XX strain used to create *sdc-2 (y93, RNAi) XX* embryos); TY0525, *him-8(e1489)* IV; *xol-1(y9)* X (used for XX and XO DCC bound). Strain to generate XO males lacking DCC binding: CB1489, *him-8(e1489)* IV (used for XO DCC not bound).

Sample size. No statistical methods were used to predetermine sample size.

ChIP-seq, RNA-seq and chromosome conformation capture. To obtain wild-type control embryos, wild-type N2 worms were grown at 20 °C on NG agar plates with concentrated HB101 bacteria. For DC mutant embryos, 10 μl of packed synchronous *sdc-2(y93)* L1 worms were placed onto 10 cm RNAi plates (NG agar with 1 mM IPTG and 100 μg/ml Carbenicillin) seeded with 2–3 ml of concentrated HT115 (DE3) bacteria carrying the Ahringer feeding library plasmid[25] expressing the coding region of *sdc-2*. The RNAi plates were incubated at 25 °C overnight before L1 larvae were added.

Immunofluorescence and FISH analysis. Animals were grown at 20 °C on NG agar plates seeded with OP50 grown in Luria Broth (LB). The worms were grown at 20 °C until gravid adults, then dissected for their embryos and stained as described below.

Antibodies. Rat polyclonal SDC-3 (PEM4A) antibodies were made against amino acids 1067-1340 of SDC-3 fused to GST. Rabbit polyclonal antibodies against DPY-27 (rb699) and SDC-3 (rb1079) were as described previously[18,26]. Mouse monoclonal Mab414 antibody (1 mg ml^{-1}) was obtained from Abcam (ab24609). Normal rabbit IgG (400 μg ml^{-1}) was from Santa Cruz Biotechnology (sc-2027). Rabbit polyclonal LMN-1 antibody (500 mg ml^{-1}) was from SDIX (3853.00.02)

ChIP-seq library creation and analysis. Libraries were made and analysed from one batch of wild-type embryos (data consistent with all previously wild-type published ChIP-seq data[20]) and two biological replicates of *sdc-2(y93, RNAi)* embryos as described previously[20].

Modified Hi-C embryo isolation and crosslinking. Worms of appropriate geno-type, either wild-type worms (two biological replicates) or *sdc-2(y93, RNAi)* (two biological replicates), were grown until gravid adults. The worms were collected and bleached to release the embryos and remove the carcasses. Following bleaching, embryos were centrifuged for ~45 s at 1,500–1,800 rpm and washed 3 times in 1× M9 buffer to remove bleach solution. An equal volume of 1× M9 was added to the embryos and they were frozen in 1 ml aliquots and stored at −80 °C. The frozen embryos were thawed on ice and supplemented with 1 mM PMSF and 5 mM DTT. The embryos were then washed once in 50 ml formaldehyde solution (1× M9 solution with 2% (v/v) formaldehyde, Polysciences 18814-20). Embryos were cross-linked in 50 ml of formaldehyde solution for 30 min at room temperature while shaking. Following crosslinking, embryos were washed once with 50 ml of 100 mM Tris-HCl, pH 7.5, followed by two 50 ml washes of 1× M9. The embryos were then washed once in lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl and 0.2% (v/v) Igepal CA-630 (Sigma I8896)) supplemented with 5 mM DTT, 1 mM PMSF, 0.1% (v/v) protease inhibitors (EMD 539134) and 0.5 mM EGTA. To obtain extract, embryos were dounced 10 times using the large pestle (Kontes 2 ml glass dounce, Spectrum 985-44182; clearance 0.076–0.127 mm), and then 10 times using the small pestle (clearance 0.01–0.069 mm). All douncing steps were performed on ice. The dounced extract was spun for 5 min at 100*g* at 4 °C, and the supernatant was saved. The pellet was resuspended in 750 μl of supplemented lysis buffer and dounced again. This procedure was repeated 7–10 times. After each spin, a 9 μl aliquot was taken from the supernatant, mixed with 1 μl of 10 ng ml^{-1} DAPI and visualized under a microscope. All supernatants containing only nuclei, and not broken carcasses, were combined. An aliquot of the combined supernatant was stained with DAPI and the nuclei were counted using a haemocytometer, and then spun down for 5 min at 2,000*g* at 4 °C. The nuclei were resuspended in the appropriate volume of 1.25× DpnII buffer (NEB B0543S) to create a Hi-C library as described below.

Modified Hi-C library preparation. The Hi-C libraries were made as described below. The protocol was based on a 3C library preparation followed by modifications[12,27,28]. Approximately 1.5×10^8 *C. elegans* nuclei were pipetted into 5–10 1.7 ml tubes and resuspended in 300 μl of 1.25× DpnII buffer. 38 μl of 1% (w/v) SDS was added per tube and the tubes were incubated at 65 °C for 10 min. After the addition of 34 μl of 20% (v/v) Triton X-100, the tubes were incubated at 37 °C for 1 h, shaking at 1,000 rpm. 30 μl (1,500 U) of DpnII (NEB R0543M) were added to each tube, and they were incubated overnight at 37 °C while rocking. 26 μl of 20% (w/v) SDS was added to each tube and they were

incubated at 65 °C for 20 min, shaking at 1,000 rpm. The reaction was then added to 7.6 ml of ligation master mix (745 μl of 10% Triton X-100, 745 μl of 10X T4 ligation buffer (500 mM Tris-HCl, pH 7.5, 100 mM MgCl₂, 100 mM DTT), 80 μl of 10 mg ml^{-1} BSA, 80 μl of 100 mM ATP, and 5.96 ml water). 100 μl (100 U) T4 DNA ligase (Invitrogen 15224-025) was added and the reactions were incubated for 4 h at 16 °C. After incubation 50 μl of 10 mg ml^{-1} proteinase K was added and the tubes were further incubated at 65 °C overnight. The next day, 50 μl of 10 mg ml^{-1} proteinase K was added to the reactions, and they were incubated at 65 °C for an additional 2 h. 2 μl of RNaseA (1 mg ml^{-1}) was added to each sample and incubated for 30 min at 37 °C. The ligated DNA was then phenol-chloroform extracted and ethanol precipitated overnight. DNA was pelleted at 14,000*g* for 30 min at 4 °C, and then washed twice with 70% ethanol and air-dried. The DNA pellets from all Hi-C reactions were combined and dissolved in a total of 500 μl of 1× TE buffer, pH 8.0. Excess salt was removed from the samples via centrifugation using a filter unit (AMICON Ultra Centrifugal Filter Unit – 0.5 ml 30 kDa) following the manufacture's instruction. Briefly, the samples were spun at 18,000*g* for 10 min to reduce the volume to 40–50 μl. Flow through was discarded and 450 μl of 1× TE, pH 8.0 buffer was added to each unit and spun as before. This wash step was repeated at least 5 times. The volume of the eluate was adjusted to 100 μl with water. The concentration of DNA was determined and 10 μg of the Hi-C library was resuspended in 100 μl of water. AMPure beads, supplied as a suspension of magnetic beads in a PEG solution (Beckman Coulter, A63880), were used to remove large DNA fragments (>10 kb), following the protocol provided by the manufacturer. Specifically, for the first DNA selection, 35 μl of AMPure beads were added to the 100 μl of DNA. The supernatant was kept and the beads, which bind only large DNA molecules under these PEG conditions, were discarded. To then remove smaller fragments, 65 μl of AMPure beads were added to the supernatant and the beads, which bind all DNA molecules greater than 100 bp due to the greater PEG concentration, were kept and washed with 70% ethanol. The DNA was eluted from the beads in 100 μl of 10 mM Tris-HCl, pH 8.5. The eluted DNA was then adjusted to 125 μl with 1× TE, pH 8.0 and sheared to 500–1,000 bp using a Covaris S2 (Covaris, 520045) in micro tubes with the following settings: duty cycle, 5%; intensity, 3; cycles/burst, 200; time, 65 s. The sheared DNA was then size selected for fragments larger than ~100 bp using AMPure beads and eluted in 34 μl of water. The DNA was quantified and 500 ng was used to make a paired-end Illumina sequencing library following the standard protocol (PE-930–1001), with the exception that we size selected 500–600 bp at the gel excision step before adding adapters for sequencing. The library was sequenced using 100 bp paired end reads with a HiSeq2500 s machine.

Read mapping/binning/ICE correction. Iterative mapping and error correction of the chromatin interaction data were performed as previously described[29]. Supplementary Table 1 summarizes the mapping results and lists the different categories of DNA molecules encountered in the libraries. We obtained around 70 million valid pairs that represent chromatin interactions per replicate. The frequency of redundant read pairs, due to PCR amplification were found to be below ~5% and were removed. The number of Hi-C interactions mapped to sequences belonging to homologous chromosomes (both intra-chromosomal (*cis*) and inter-homologue (*trans*) interactions) was much higher than the interactions mapped to non-homologous chromosomes (inter-chromosomal (*trans*) interactions). Assuming that inter-homologue interactions (*trans*) are as frequent as non-homologous inter-chromosomal interactions (*trans*), we estimate that 80–90% of interactions mapped to the same chromosomes are intra-chromosomal (*cis*) interactions, with DC mutants (90%) higher than wild type (>85%). Whether this difference reflects a biological phenomenon or is due to technical differences is currently not known. Conversion of interaction data into *Z*-scores eliminates this difference (see below).

The data were binned at both 10 kb and 50 kb non-overlapping genomic intervals. Binned data were normalized for intrinsic biases such as differences in number of restriction fragments within bins using the previously developed ICE method[29]. To normalize for differences in read depth of different data sets we summed the entire genome-wide binned ICE-corrected interaction matrix, excluding the diagonal ($x = y$) bins. We then transformed each interaction into a fraction of the matrix sum (minus diagonal $x = y$ bins). Each fraction was then multiplied by 10^6. Biological replicates were highly correlated (Pearson's correlation coefficients >0.98 for 50 kb binned data excluding short-range interactions up to 50 kb). The correlations between biological replicates were higher than those between the wild type and DC mutant. Overall these numbers indicate that the modified Hi-C procedure was reproducible and performed as expected. For most analyses sequence reads obtained for biological replicates were pooled and ICE-corrected as described above to create a combined replicate data set.

At 10 kb resolution, very long-range interactions are not sampled deeply enough to provide robust and reliable data. Therefore, we truncated the 10 kb binned data to include only *cis* interaction pairs separated by 4 Mb or less in linear genomic distance. This distance cutoff was chosen based on the observation that

beyond this point, both wild-type and DC mutant data sets have no observed reads in more than 50% of bin–bin interactions. In addition to limiting the dynamic range of interaction counts at these large distances, this high frequency of un-sampled interactions beyond 4 Mb causes a dramatic collapse in the standard deviation of the overall chromatin interaction decay over distance, making the LOWESS expected and Z-score calculations beyond 4 Mb unreliable. For 50 kb bins, all distances were included in analyses, because the coverage of *cis* interaction pairs never dropped below 50% for any distance at this resolution.

TAD calling (insulation square analysis). To calculate the 'insulation' score of each bin in the 10 kb binned Hi-C data, we calculated the average number of interactions that occurred across each bin. This can be visualized by sliding a 500 kb × 500 kb (50 bins × 50 bins) (Extended Data Figs 2 and 3) square along the matrix diagonal, and aggregating all signal within the square. The mean signal within the square was then assigned to the 10 kb diagonal bin and this procedure was then repeated for all 10 kb diagonal bins. For any bins within 500 kb of the matrix start/end, an insulation score was not assigned, as the 500 kb × 500 kb insulation square would extend beyond the matrix bounds. The insulation score was then normalized relative to all of the insulation scores across each chromosome by calculating the $\log_2$ ratio of each bin's insulation score and the mean of all insulation scores. Valleys/minima along the normalized insulation score vector represent loci of reduced Hi-C interactions that occur across the bin. These valleys/minima are interpreted as TAD boundaries or areas of high local insulation. The valleys/minima were detected as follows: first, a delta vector was calculated to approximate the slope of the normalized insulation vector. The delta vector is defined as the difference between the amount of insulation change 100 kb to the left of the central bin and 100 kb to the right of the central bin (relative to the central bin) (Extended Data Fig. 3a, b). The delta vector crosses the horizontal 0 at all peaks and all valleys. All bins where the delta vector crosses 0 were extracted. Zero-crossings occurring at peaks were removed, and the remaining zero-crossings, all occurring at potential valleys were passed through a boundary strength filter. The boundary strength was defined as the difference in the delta vector between the local maximum to the left and local minimum to the right of the boundary bin. All boundaries with a boundary strength <0.1 were removed. This method in practice is very similar to the widely used zero-derivative method for detecting peaks/valleys in various signal vectors.

The precision with which we could define a boundary was determined by comparing boundary calls across biological replicates (Extended Data Fig. 3c). The final boundary zones were defined as ±30 kb around the pooled replicate insulation minima bins (70 kb total) because most (>80%) replicate boundary calls overlapped within this window. Wild-type and DC mutant insulation profiles were compared by subtracting the wild-type insulation profile from the DC mutant insulation profile. We compared the insulation profiles and boundary calls resulting from a full range of alternative insulation square sizes (Extended Data Fig. 2b, c). We find that a 500 kb square size captures best the major robust boundaries that change in the DC mutant. In contrast, boundaries detected by a 100 kb insulation square, for example, only affect interactions within a few bins of the boundary rather than insulating larger genomic regions from one another and do not change in the DC mutant (Extended Data Fig. 2e).

Code availability

Code for Hi-C read mapping and processing is based on the published ICE method[29]. The code to calculate insulation profiles is publicly available at (https://github.com/blajoie/crane-nature-2015).

Z-score calculation. We modelled the overall chromatin interaction decay with distance with a modified LOWESS method (alpha = 0.5%, ignore zeros, IQR filter), as described previously[30]. LOWESS calculates the weighted-average and weighted-standard deviation for every genomic distance by leveraging all data genome-wide. We transformed interaction data into a Z-score by calculating: ((observed signal – LOWESS-average)/LOWESS-stdev). Observed signals with a count of 0 were excluded from the Z-score transformation. By expressing inter-action data as Z-scores, we corrected for minor differences in the overall decay with genomic distance that can vary slightly between samples.

To calculate the difference between the wild-type and DC mutant Hi-C data, we calculated the difference between the combined replicate DC mutant Z-score data and the combined replicate wild-type Z-score data (DC mutant Z-score minus wild-type Z-score). (Fig. 1c, f and Extended Data Figs 1, 4–6).

Compartment analysis and comparison to LEM-2 associated domains. The presence and locations of A/B-compartments can be quantified using principle component analysis, where the largest eigenvector typically represents the com-partment profile[11,13,29]. Applying this approach to 50 kb binned interaction data, we determined the positions of such preferentially associating compartments along each *C. elegans* chromosome (Extended Data Figs 4e, 5e and 6c, g, k, o). Compartment positions quantified in this manner closely align with the large sub-chromosomal domains that are visible in the chromatin interaction maps.

LEM-2 binding data[15] ($\log_2$ ratio of ChIP signal over input) were lifted from the ce4 genome assembly to the ce10 assembly, and data were averaged in 50 kb bins. These bins correspond exactly to the coordinates of the binned chromatin inter-action data. Binned LEM-2 binding data were then plotted along each chro-mosome, and compared to the compartment profiles (Extended Data Figs 4e; 5e and 6c, g, k, o).

3D plots. To test for elevated levels of interaction between certain classes of sites in the genome, we constructed 3D plots. For each plot, a list was first made of all 10 kb bins meeting desired criteria: containing any *rex* or predicted *rex* (*Prex*) site (Fig. 3d), containing a *rex* or *Prex* site in the top 25 by ChIP-seq signal (Fig. 3d and Extended Data Fig. 7f), or containing any *dox* site (Extended Data Fig. 7g). *Prex* sites are defined as those with very strong ChIP-seq signal that was greatly dimin-ished in *sdc-2* mutants. Unlike *rex* sites, which also have these properties, *Prex* sites have not been tested for autonomous DCC recruitment *in vivo* through an array assay[4]. Next, sub-matrices of wild-type or DC mutant interactions were prepared for all possible pairs of bins in this list, extending 50 kb away from the central bin in all directions. Pairs of bins that were separated by less than 100 kb were excluded so that no sub-matrices would overlap the whole-chromosome interaction matrix diagonal (interactions within the same bin). All pairwise sub-matrices were then averaged together and the values plotted in 3D. If sub-matrices stretched past the end of the chromosome or overlapped bins with no data (unmappable sequence, etc.), only the part of the sub-matrix containing data was included in the average.

Cumulative plot randomization. To assess the significance of the decrease in Z-scores observed for the set of *rex–rex* interactions, we selected 1,000 random sets of 785 interactions (Fig. 3c and Extended Data Fig. 7e). These random interaction sets were thus the same size as the *rex–rex* interaction set. The P value represents the fraction of the 1,000 randomized interaction sets that changed more from wild-type to DC mutant than the *rex–rex* set (according to the KS test statistic).

Circos plots. Plots were generated using the Circos package to highlight the strength of various sets of *rex–rex* interactions in wild-type and DC mutant at 50 kb resolution. A Z-score threshold of 2 was selected and interactions were colored and given line thickness proportional to their Z-score. Z-scores greater than 8 were determined to correspond to 'singleton' outlier interactions and were excluded.

TAD FISH

Preparation of FISH probes. FISH probes covering 400–500 kb genomic regions were prepared using pooled fosmids (BioScience LifeSciences), as described previously[8]. 1 µg DNA was labelled with Alexa-488, Alexa-594, Alexa-555 or Alexa-647 using FISH Tag DNA Kit (Invitrogen). The genomic locations of tested regions are listed as follows: Probe1, chromosome X, 9.05–9.45 Mb; Probe2, chro-mosome X, 9.5–9.9 Mb; Probe3, chromosome X, 9.95–10.35 Mb; Probe4, chromosome X, 2.0–2.5 Mb; Probe5, chromosome X, 2.5–3.0 Mb; Probe6, chromosome X, 3.0–3.5 Mb; Probe7, chromosome X, 11.2–11.7 Mb; Probe8, chro-mosome X, 11.7–12.3 Mb; Probe9, chromosome X, 12.3–12.8 Mb; Probe10, chro-mosome X, 10.6–11.1 Mb; Probe11, chromosome X, 11.1–11.6 Mb; Probe12, chromosome I, 4.1–4.6 Mb; Probe13, chromosome I, 4.6–5.1 Mb; Probe14, chro-mosome I, 5.1–5.6 Mb; and Probe15, chromosome X, 3.5–4.1 Mb

FISH procedure. *C. elegans* embryos were obtained by dissecting gravid N2, *him-8(e1489)* or *szT1/sdc-2(y74) unc-3(e151)* adults in 13 µl of water on poly-lysine coated slides. A coverslip was added on top of the dissected worms, and the slides were then frozen in liquid nitrogen for at least 1 min. Coverslips were cracked off, and the samples were dehydrated in 95% ethanol for at least 10 min. 35 µl of fix (2% (v/v) paraformaldehyde in egg buffer (25 mM HEPES, pH 7.3, 118 mM NaCl, 48 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$) was added and slides were incubated in a humid chamber for 5.5 min. Slides were washed 3 times for 10 min with 1× PBS-T (0.5% Triton X-100 in 1× PBS) at room temperature. Excess 1× PBS-T was then removed and 15 µl of hybridization solution (30% (v/v) formamide, 3× SSC, 10% dextran sulphate) containing approximately 50 ng of each FISH probe were added. Hybridization was performed in a temperature-controlled slide chamber (Bio-Rad ALD0211 Alpha Unit Block Assembly). The following FISH program was typically run overnight: 90 °C for 5 min, 0.5 °C per second to 50 °C, 50 °C for 1 min, 0.5 °C per second to 45 °C, 45 °C for 1 min, 0.5 °C per second to 40 °C, 40 °C for 1 min, 0.5 °C per second to 38 °C, 38 °C for 1 min, 0.5 °C per second to 37 °C, 37 °C overnight. Slides were then washed at 39 °C as follows: 3 times for 10 min with 30% (v/v) formamide in 2× SSC, 3 times for 10 min with 20% (v/v) formamide in 2× SSC, 3 times for 5 min with 10% (v/v) formamide in 2× SSC, 3 times for 5 min with 2× SSC, and 3 times for 1 min with 1× SSC. Slides were then washed 3 times for 10 min in 1× PBS-T. For N2 embryos, the slides were mounted in Prolong Gold antifade reagent (Invitrogen, P36934) containing DAPI (1 ng ml^{-1}). For *him-8(e1489)* and *sdc-2(y74)* embryos, immunostaining with SDC-3 antibody was performed following FISH to determine the sex and/or genotype of embryos as described below.

Immunofluorescence. Excess 1× PBS-T was removed and 35 µl of primary antibody (rat anti-SDC-3 antibody, 1:400) were added. Samples were incubated in a humid chamber for 6 h to overnight. Slides were washed 3 times for 10 min with 1× PBS-T at room temperature and then incubated in secondary antibody (Alexa-Fluor-647 goat anti-rat antibody (Invitrogen), 1:250) for 6 h to overnight. Slides were then washed 3 times for 10 min with 1× PBS-T at room temperature and then mounted.

Microscopy and co-localization analysis. Embryos were imaged on a Leica TCS SP8 microscope using 63×, 1.4 NA objective lenses. The scanning settings for SP8 were: 1,024 × 1,024 pixels frame size, 51.5 nm pixel size, 3.5 zoom factor, 400 Hz scanning speed and 83.9 nm step size for z sections. Image deconvolution was performed using Huygens Professional Software.

After deconvolution, the homozygous *sdc-2(y74) unc-3(e151)* XX embryos were determined based on the lack of SDC-3 staining on the X chromosomes and their sex was further confirmed by examining the number of X-chromosome FISH signals. For all genotypes, embryos between 200-cell and 400-cell stages which match the developmental stage of Hi-C samples were selected for further analysis.

The deconvolved image stacks of embryos were manually segmented based on DAPI staining using Priism software[31]. FISH signals in individual embryos were thresholded to make the total signals from each probe occupy equal volume. The centre-of-mass coordinates for the FISH signals from the probe in the middle of the probe set were determined using a built-in find points function in Priism. Regions of equal volume were then created around the FISH signals to encompass the entire sets of FISH signals on the same chromosomes using a Python script. Pearson's correlation coefficients between pairs of FISH probes were then calculated: the more the two probes overlap, the higher the correlation coefficient.

3D quantitative FISH for measuring the interaction frequency between genomic loci

FISH experimental design. To examine the DCC dependence of interactions between genomic loci, and to distinguish between inter-homologue (*trans*) and intra-chromosomal (*cis*) interactions, we performed the 3D FISH analysis in both XX and the XO embryos in which the DCC was bound or not bound to X chromosomes. For these experiments, we acquired confocal images of embryos hybridized with FISH probes to two genomic loci and also stained with lamin (LMN-1) antibody and DAPI to help segment the nuclei. Newly developed software was used to measure the 3D distance between FISH probes automatically.

To assay XO embryos having DCC binding on the X chromosome, we performed the experiments using *xol-1(y9); him-8(e1489)* animals. These animals carried a deletion of the master switch gene (*xol-1*) that inhibits DCC binding to X chromosomes of XO embryos. DCC association with the X chromosome kills XO animals by the L1 larval stage. To enrich for XO male embryos in our experiments, we used mutation in *him-8* (high incidence of males), which elevated the frequency of male progeny in a hermaphrodite brood from 0.02% to 37%. The XX embryos deficient in DCC binding were obtained from *szT1/sdc-2(y74) unc-3(e151)* animals, as described above.

To measure the distance between FISH foci in z stacks of confocal images, we developed software (Mets and Meyer, unpublished) that identified foci automatically, assigned foci to appropriate nuclei, and quantified the distance between foci in 3D space, thereby permitting the unbiased quantification of probe-interaction frequency. The quantification involved several steps. Each FISH spot was centre fitted, and its location was recorded in x, y and z. For all nuclei, distances between all combinations of red and green FISH spots were calculated using a distance quantification algorithm that employs LMN-1 and DAPI co-staining to segment the nuclei. In XX embryos, four FISH spots (two red and two green) were generally apparent for X-linked probes in each nucleus, corresponding to the hybridization of both probes to their target sites on both homologous chromosomes. To eliminate the bias in our calculations for interactions caused by the inclusion of distances between probes on different chromosomes, we used only the shortest of the four possible distances between red and green probes in each nucleus for X-linked loci in XX embryos and for autosomal loci in all embryos.

We segmented the distances into 300 nm bins and plotted the relative contribution of each bin to the total number of measured distances. The limit of resolution of the confocal microscope is ~200 nm in x and y, making 300 nm a reasonable choice for the smallest bin. Furthermore, probes spaced <260 nm apart appear overlapping by visual inspection, and probes spaced ~700 nm apart appear adjacent, indicating that the smallest bin size (300 nm) represents a degree of overlap that would be consider co-localized. Chi-square tests comparing the number of FISH pairs within 0–300 nm to those within 301–2,700 nm were used to assess the similarity of data sets from different classes of embryos. The unbinned data were also represented in cumulative plots (Extended Data Fig. 9a–f).

Preparation of FISH probes. Primers were created to amplify 3–6 kb sequences of DNA corresponding to each site. 1 µg of the probe DNA was labelled using the FISH tag DNA Red Kit (Molecular Probes, F32949) or the FISH tag DNA Green

Kit (Molecular Probes, F32947) according to the manufacturer's protocol, with the following exceptions: the DNaseI was diluted 1:1,000, and the labelled probes were eluted in 10 µl but then diluted 1:10 for use in staining. Primers to make the probes are listed below: *rex-23* F (gcccattcaacccattgtcc); *rex-23* R (gcactcgcatattccaaaacg); *rex-32* (cgcagctggccgttaaatg); *rex-32* R (cattgcaggtgcgttcacaac); *rex-47* F (ccgaaa cacaacaacaatgc); *rex-47* R (agactggcgaagaggaacaa); *rex-8* F (tgtgatgcaagccagag ttgg); *rex-8* R (cattgagccgaatttccaaagg); *rex-14* F (ttgcagttgcgaaagaaatg); *rex-14* R (tttttgaggagatcgggatg); *rex-1* F (ctcaagagctgcgaagtgc); *rex-1* R (aaagttcaacgaccag aatgc); Xnb1 F (tcgaatgacctcaagcactg); Xnb1 R (tcaccactgaaatcggcata); Xnb2 F (aaaacgcggtgaaacgatac); Xnb2 R (gttttcctctccccaacaca); Xnb3 F (gtatgcacacgcctc aaaaa); Xnb3 R (ttggaatctctcaccggagt); Xnb4 F (atggtaggacgttccgtttg); Xnb4 R (aatccagccctctggtttttc); Xnb5 F (atttgcttgggcattaaacg); Xnb5 R (ttcaatgaagagacgc gatg); Xnb6 F (ccgtttttggcaatgaactt); Xnb6 R agaggatgggtttggacgttg); Xnb7 F (gagcg acgattctgtcttcc); Xnb7 R (cgtcatgtccattttgcttg); Xnb8 F (atcgtgccaagacctattcg); Xnb8 R (ttttcgcatttcctgcttct); Inb1 F (aaaggaccctcccccctaact); Inb1 R (tccatgccta cttgcctacc); Inb2 F (caggcgagcattctaccact); Inb2 R (ccggaaagagcattgattgt); Inb3 F (gcactgcaattgccaaccag); Inb3 R (ttcaaagacactcctcccatcc); Inb4 F (attgccgctaacc aagtgc); and Inb4 R (tccaacgccaacaaaactcc).

Combined FISH and immunofluorescence procedure. FISH followed by immunofluorescence was performed as described in the previous section. 5–10 ng (0.5–1 µl of 1:10 dilution) of each FISH probe was used for hybridization. For immunofluorescence, primary antibodies were applied at the following dilutions in 1× PBS-T: rat anti-SDC-3, 1:400; rabbit LMN-1, 1:400. Secondary Alexa-Fluor-555 donkey anti-rabbit and Alexa-Fluor-647 donkey anti-rat antibodies (Invitrogen) were used at a 1:200 dilution.

Microscopy and image analysis. Embryos were imaged on a Leica TCS SP2 AOBS confocal microscope or a Leica TCS SP8 microscope using 63×, 1.4 NA objective lenses. The scanning settings for SP2 were: 1,024 × 1,024 pixels frame size, 46.5 nm pixel size, 5.0 zoom factor, 400 Hz scanning speed and 81 nm step size for z sections. The scanning settings for SP8 were as described in the previous section. The images were then deconvolved using Huygens Professional with the appropriate settings. The images were visualized and processed in Priism. The embryos were first cut out from the background using the edit polygon and cut mask function. Then the DAPI and LMN-1 channels were blurred using the 3D Filter Function to make the nuclear signal continuous and thus allow for the nuclei to be accurately segmented. This protocol permits each nucleus to be counted as one spot by the find points function. A new processed image was made by discarding the z sections in the top and bottom 10% of the image, and by substituting the new blurred channels for those in the original image. The find points function was then used to count and record the local centre of mass (LCOM) of each nucleus and each FISH spot in x, y and z using user-defined threshold values. The data for the location of the nuclei and the FISH, along with the processed image are processed using the software described in FISH experimental design section above.

***rex-47* deletion**

Plasmid construction. Expression vectors for both codon-optimized Cas9 and sgRNA (Peft-3::cas9-SV40_NLS::tbb-2 3′ UTR and PU6::unc-119_sgRNA[32] were obtained from Addgene. To enhance the expression and assembly of sgRNA, the sgRNA vector was modified by introducing an A-U flip in the sgRNA stem loop and extending the Cas9 binding hairpin[33]. To clone the protospacer sequence for the sgRNA targeting *rex-47* (5′-GTAGTCACACCGAATTGATA-3′), the modified sgRNA vector was PCR amplified using primers GTAGTCACACCGAAT TGATAGTTTAAGAGCTATGCTGGAAACAGCATAG and AACAGCTATG ACCATGATTACGCCAAGCTTCACAGCCGACTATGTTTGGCGTCGAG or GACGTTGTAAAACGACGGCCAGTGAATTCCTCCAAGAACTCGTACAAA AATGCTCTGAAG and TATCAATTCGGTGTGACTACAAACATTTAGATT TGCAATTCAATTATATAG to generate two fragments with overlapping protospacer sequences. The two PCR products were then inserted into the sgRNA vector backbone generated by EcoRI/HindIII digestion using a previously described Gibson Assembly protocol[34]. To clone the repair template for making the 419 bp *rex-47* deletion, two 500 bp homology arms flanking the target region were PCR amplified from *C. elegans* genomic DNA using primers ACGACG TTGTAAAACGACGGCCAGTGAATTCGACGTGTCGAAATTTTTCAG and TTGAATTATTGACCATGGCAGACAGAGCGTAACGAGTAAT or ACGC TCTGTCTGCCATGGTCAATAATTCAATGCAATGAAG and CTATGACC ATGATTACGCCAAGCTTAATAATAAACTTCCATAAGA. The homology arms and the sgRNA vector backbone were assembled using Gibson Assembly. The resulting repair template contains an NcoI restriction site between the homology arms, which facilitates the identification of desired mutations.

Cas9-mediated mutagenesis and mutant screening. To generate Cas9-mediated heritable *rex-47* deletion, DNA microinjection was performed according to standard protocols. The Cas9 expression vector, sgRNA expression vector, repair

template and two co-injection markers: pCFJ90 (Pmyo-2::mCherry) and pCFJ104 (Pmyo-3::mCherry) were mixed and injected into the germline of 34 N2 young adults at the following concentrations: Cas9 (50 ng μl^{-1}), sgRNA (200 ng μl^{-1}), repair template (50 ng μl^{-1}), pCFJ90 (2.5 ng μl^{-1}) and pCFJ104 (5 ng μl^{-1}). Three days post-injection, 269 F1 s expressing both Pmyo-2::mCherry and Pmyo-3::mCherry markers were cloned into liquid culture in 96-well plates and propagated at 20 °C as described previously[35]. Worms from each well were lysed and PCR amplified using primers CCGAAACACAACAACAATGC and TGGTA GCCGTATGCACAGTT. We identified 8 deletion mutants from the 269 F1s (3%) based on the size of PCR products. These deletions were further verified by NcoI digestion of the PCR fragments. The progeny of the F1s carrying the *rex-47* deletion alleles were then cloned into a new set of wells for the identification of homozygote mutants. PCR products from the homozygote mutants were sequenced to verify the precision of the deletions.

ChIP–qPCR. Wild-type and *rex-47* deletion embryos were obtained as described earlier. Input and ChIP samples using rabbit anti-DPY-27 or rabbit anti-SDC-3 antibody were prepared according to previously published protocols[20]. Three pairs of qPCR primers (ACTTTGCAAGAGTATGTAGTGAA/ACGAGTAATACTT TGAGCATACTT, TACGGCTACCAATCTTGTAA/TCTGTATCTCTAATCC CTAATAGT and TGTGACTACTTGCCCAATAAA/TATCTCTCCCTTCGCC TAAA) were used to amplify three ~100 bp regions located upstream, downstream or within the *rex-47* deletion region, respectively. qPCR was performed using iQ SYBR Green Supermix (Bio-Rad,170-8880) on a CFX384 Touch Real-Time PCR Detection System (Bio-Rad).

FISH analysis of rex-47 deletion strain. The legend for Fig. 3h provides the quantification for three-way comparisons of FISH probe colocalization among wild-type, DC mutant, and *rex-47* deletion strains. For two-way comparisons using the one-tailed Mann–Whitney U-test, the *rex-47* deletion strain differed significantly from the wild-type strain ($P < 10^{-5}$) for probes on each side of the TAD boundary, and the *rex-47* deletion strain was not statistically different from the DC mutant strain ($P = $ NS), as expected.

RNA-seq library creation. Embryos of appropriate genotype, four total wild-type biological replicates (two from the Hi-C biological replicates) and three total *sdc-2* (*y93*, RNAi) biological replicates (two from the Hi-C biological replicates), were isolated following the procedures above and frozen at −80 °C in 1× M9 buffer. RNA was extracted using a protocol described previously[36], except that 10 μl of a 20 mg ml^{-1} glycogen solution was used as a carrier. Libraries were prepared from 10 μg of total RNA. PolyA RNA was purified using the Dynabeads mRNA purification kit (Ambion) and fragmented using Fragmentation Reagent (Ambion). First strand cDNA was synthesized from polyA RNA using the SuperScript III Reverse Transcriptase Kit with random primers (Life Technologies). Second strand cDNA synthesis was performed using Second Strand Synthesis buffer, DNA Pol I, and RNase H (Life Technologies). cDNA libraries were prepared for sequencing using the mRNA TruSeq protocol (Illumina).

Gene expression analysis. Libraries were sequenced with Illumina's HiSeq2000 platform. Reads were required to have passed the CASAVA 1.8 quality filtering to be considered further. To remove and trim reads containing the sequencing barcodes, we used cutadapt version 0.9.5 (http://code.google.com/p/cutadapt/). Reads were aligned to the transcriptome using GSNAP[37] version 2012-01-11. Uniquely mapping reads were assigned to genes using HTSeq version 0.5.4p3 using the union mode. Gene expression levels and changes in gene expression were determined by analysis with DESeq[38]. Gene expression analysis were conducted both with these RNA-seq data sets and published GRO-seq data sets[20]. For each chromosome, scatter plots analysed the $\log_2$ of the median fold-change in gene expression (DC-mutant expression/wild-type expression) calculated for each 10 kb bin along the chromosome versus the change in insulation score for that bin in wild-type versus DC mutant embryos. No significant correlation was found between the change in gene expression and the change in insulation score: for chromosomes I, II and X, $R = 0.04$; for chromosome III and IV, $R = 0.00$; for chromosome V, $R = 0.03$.

25. Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* **30**, 313–321 (2003).
26. Chuang, P. T., Albertson, D. G. & Meyer, B. J. DPY-27:a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**, 459–474 (1994).
27. van Berkum, N. L. & Dekker, J. Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol. Biol.* **567**, 189–213 (2009).
28. Belton, J. M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
29. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9**, 999–1003 (2012).
30. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
31. Chen, H., Hughes, D. D., Chan, T. A., Sedat, J. W. & Agard, D. A. IVE (Image Visualization Environment): a software platform for all three-dimensional microscopy applications. *J. Struct. Biol.* **116**, 56–60 (1996).
32. Friedland, A. E. *et al.* Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods* **10**, 741–743 (2013).
33. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
34. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).
35. Lo, T. W. *et al.* Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics* **195**, 331–348 (2013).
36. Baugh, L. R., Demodena, J. & Sternberg, P. W. RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* **324**, 92–94 (2009).
37. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
38. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
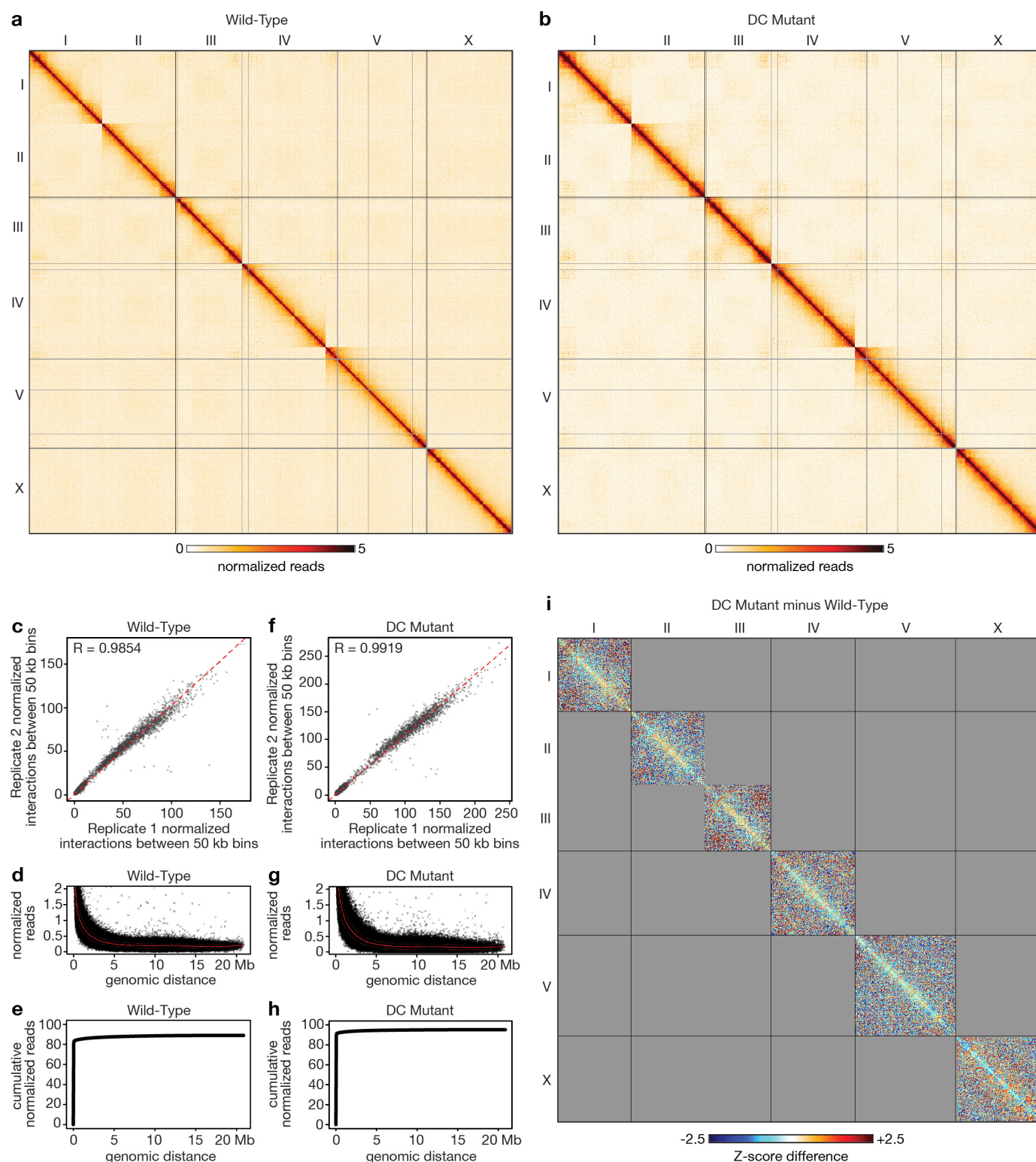
Extended Data Figure 1 | Genome-wide chromatin interaction maps for wild-type or DC mutant embryos and genome-wide difference chromatin interaction map. a, b, Genome-wide chromatin interaction maps for wild-type embryos (**a**) and DC mutant embryos (**b**) from Hi-C data of two biological replicates pooled and binned at 50 kb and corrected with ICE. **c, f,** Scatter plots comparing normalized interactions between pairs of 50 kb bins in the two biological replicates from wild-type embryos (**c**) or DC mutant embryos (**f**) (both excluding $x = y$ diagonal). A strong correlation between biological replicates is shown for wild-type embryos (Pearson's correlation coefficient = 0.9854) and for DC mutant embryos (Pearson's correlation

coefficient = 0.9919). **d, g,** Overall interaction frequency decays with increasing genomic distance in wild-type embryos (**d**) and in DC mutant embryos (**g**). **e, h,** Cumulative reads versus linear genomic distance in wild-type embryos (**e**) and in DC mutant embryos (**h**). **i,** Genome-wide difference chromatin interaction map. Shown is the 50 kb binned heatmap depicting the *Z*-score difference between wild-type and DC mutant embryos (see Methods for *Z*-score difference calculation). The most apparent differences are on the X chromosome: blue signal within TADs (loss of intra-TAD interactions) and red signal between TADs (gain of inter-TAD interactions).
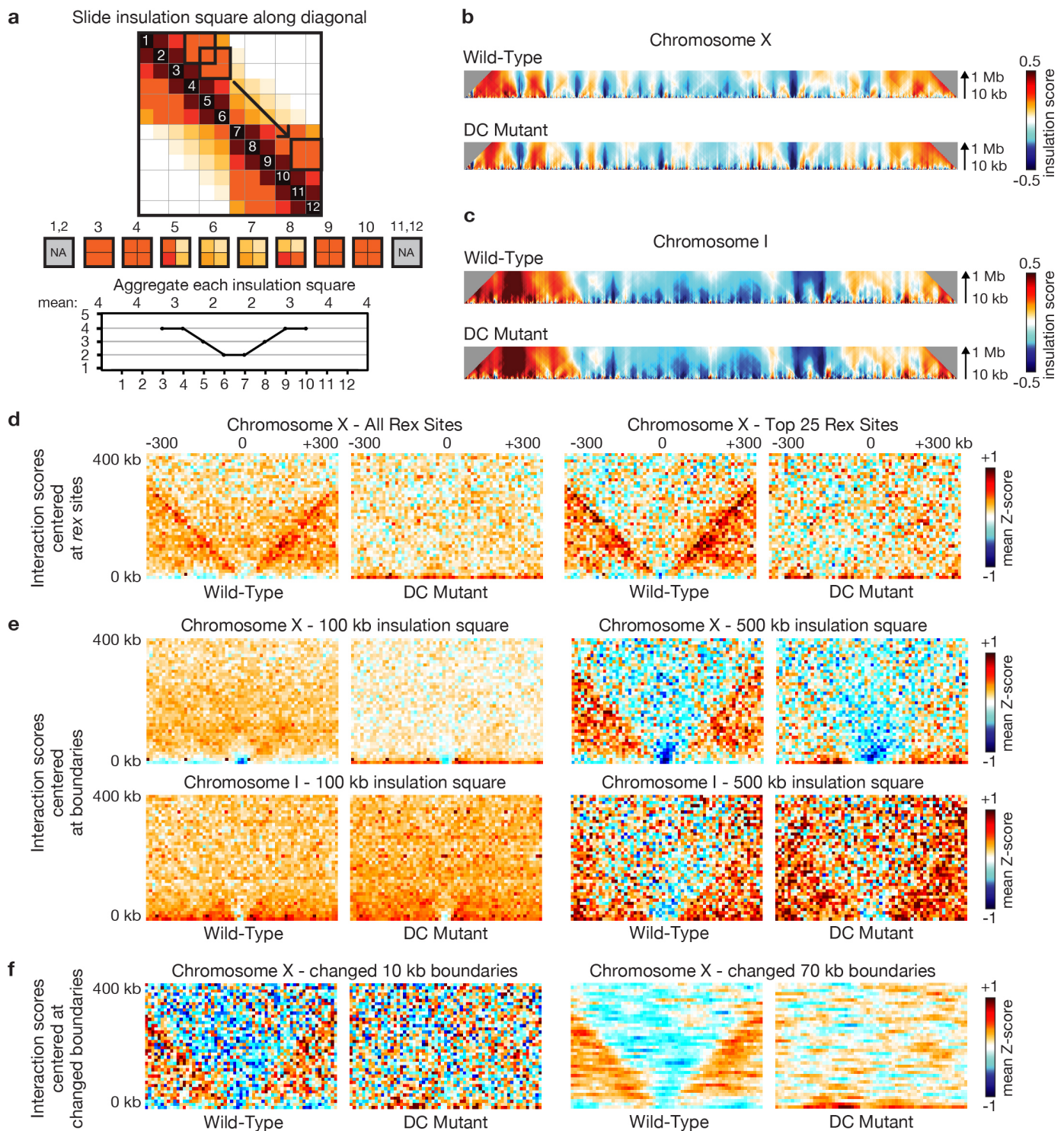
Extended Data Figure 2 | Insulation profile calculation parameters and boundary calling. a, Cartoon shows approach for calculating the insulation profile. A square is slid along each diagonal bin of the interaction matrix to aggregate the amount of interactions that occur across each bin (up to a specified distance upstream and downstream of the bin). Bins with a high insulation effect (for example, at a TAD boundary) have a low insulation score (as measured by the insulation square). Bins with low insulation or boundary activity (for example, in the middle of a TAD) have a high insulation score. Minima along the insulation profile are potential TAD boundaries. **b, c**, Heatmaps of chromosome X and chromosome I represent the insulation profiles calculated using insulation square sizes ranging from 10 kb to 1 Mb. At the 100 kb scale, weak boundaries are observed on the X chromosome and autosomes, but they are generally not changed in DC mutants. These boundaries cannot be detected at larger scales, meaning they do not insulate over distances beyond ~100 kb (see **e**). These

smaller scale structures may represent sub-TAD domains not correlated with dosage compensation. Boundaries called using a 500 kb insulation square represent TAD boundaries that define domains observed in chromosome-wide interaction maps of the X chromosome at 10 kb resolution. These boundaries are used in this paper (Fig. 1) and insulate over the larger distances defining the Mb-sized TADs. Boundaries on the X chromosome are the strongest and are DC dependent. **d–f**, Pile up plots depict aggregate (mean) Hi-C 10 kb Z-score data centred on specified 'anchors' (for example, *rex* sites, boundaries, changed boundaries). **d**, Pile up plots centred on all *rex* sites or top 25 *rex* sites in wild-type and DC mutant. **e**, Pile up plots centred on all boundaries called using insulation squares of 100 kb (left) or 500 kb (right) for chromosome X and chromosome I in wild-type and DC mutant. **f**, Pile up plots using boundaries called with a 500 kb insulation square, centred (left) on the single 10 kb bin at the midpoint of all 8 changed boundaries or (right) on all seven 10 kb bins within changed boundaries.

a, Local minima determination

b, Insulation Delta Calculation

c, Distribution of distances between replicate boundary calls: all chromosomes

d, TAD Boundary Strength on X and Autosomes; TAD Boundary Spacing on X and Autosomes
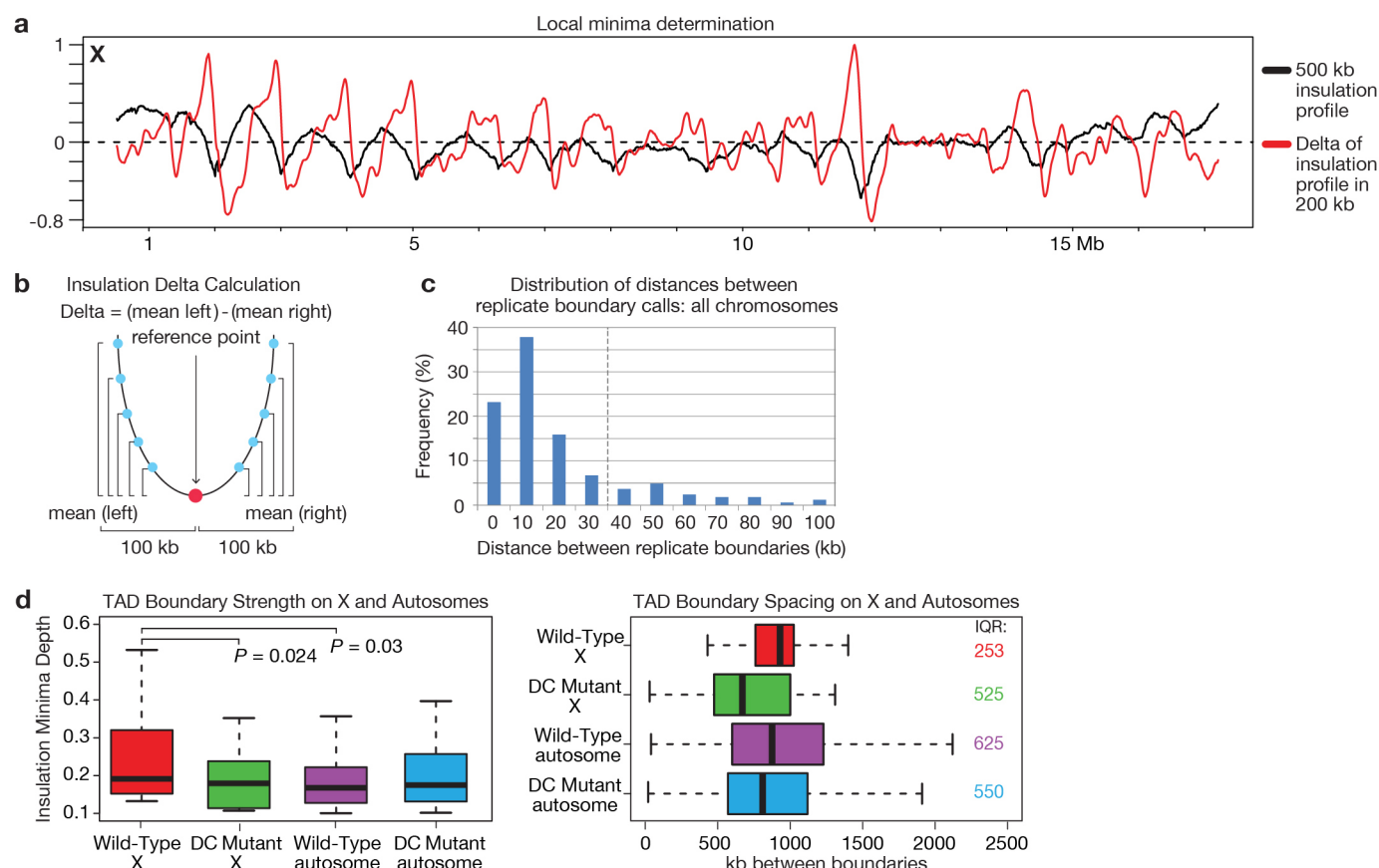
Extended Data Figure 3 | TAD boundary analysis. a, Insulation/delta plot of the 10 kb binned wild-type sample combined replicate chromosome X Hi-C data calculated using a 500-kb insulation square size. The insulation profile is depicted in black. In red, the 'delta' vector is depicted. It is derived from the insulation vector using a 200 kb delta window (see insulation methods). The 'delta' vector is used to facilitate the detection of the valleys/minima along the insulation profile. **b**, Cartoon example showing how the delta vector is calculated from the insulation data vector. For each bin (reference point) the average insulation differences are calculated between all points up to 100 kb left of the reference point relative to the reference point. The same is repeated for all points up to 100 kb right of the reference point. The delta value is then defined as the difference between the mean (left difference) and mean (right difference). **c**, Bar plot shows the distribution of distances between boundary calls obtained with biological replicate Hi-C data across all chromosomes.

Dotted vertical line indicates that ± 30 kb was chosen for boundary definition, as it was the window in which the majority of replicate boundary calls ($>80\%$) overlap. **d**, Boxplots compare boundary strength (left) and spacing (right) in wild-type versus DC mutant embryos. Wild-type boundary strength on chromosome X (defined as the distance from the insulation minimum to the largest neighbouring maximum in the insulation profile) is higher than the DC mutant chromosome X boundary strength ($P = 0.024$) and higher than the boundary strength on wild-type autosomes ($P = 0.03$). TAD boundary strength on autosomes does not change in the DC mutant compared to the wild type ($P = 0.979$). Boundaries on chromosome X have less variance in spacing (interquartile range (IQR) = 253 kb) compared to the DC mutant (IQR = 525 kb) embryos. DC mutant X chromosome boundary spacing is more similar to the boundary spacing on the autosomes in wild-type embryos (IQR = 625 kb) and DC mutant embryos (IQR = 550 kb).
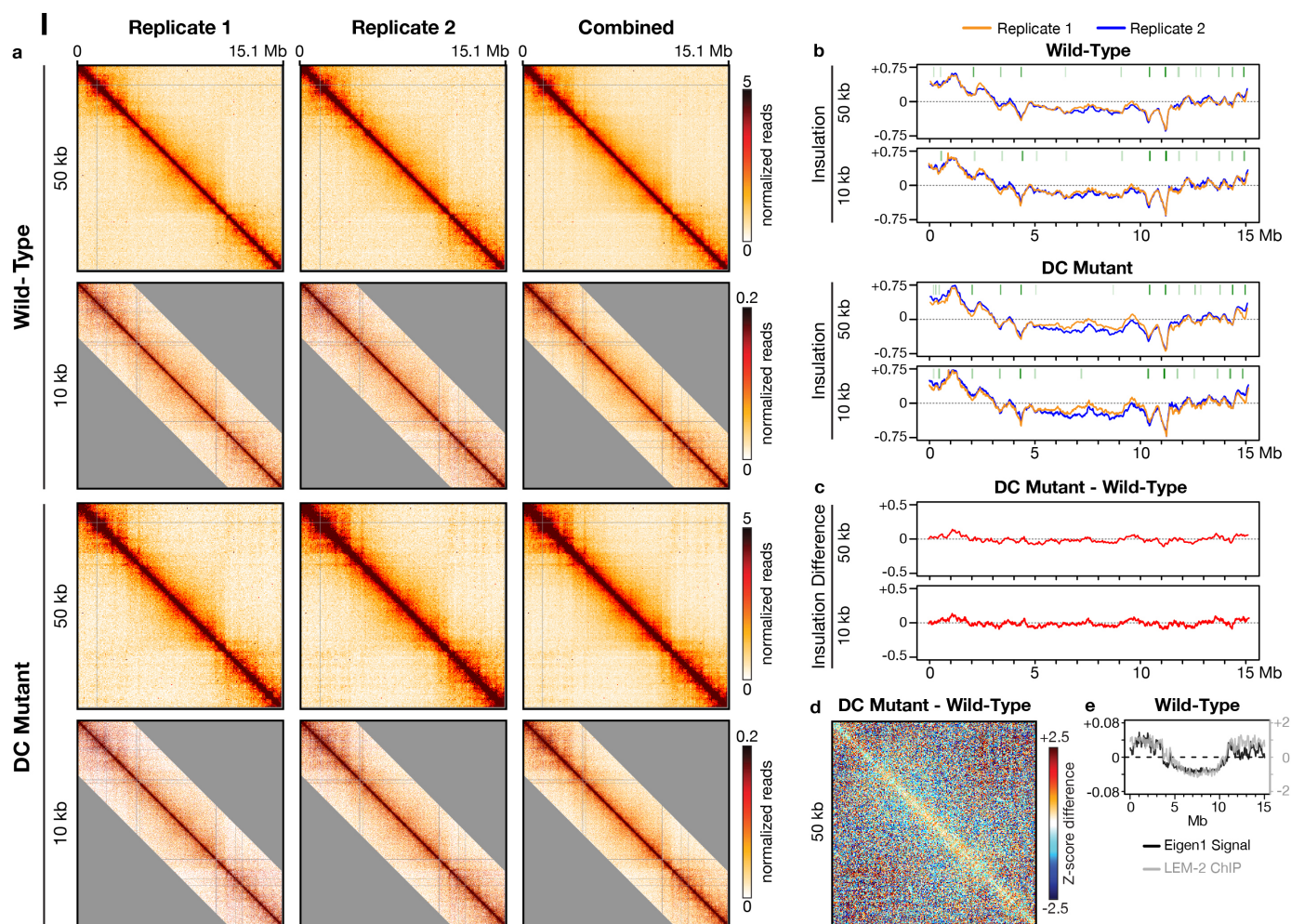
Extended Data Figure 4 | Compartment and insulation analysis for chromosome I in wild-type embryos and DC mutant embryos. **a**, ICE corrected chromatin interaction maps are shown for wild-type embryos and DC mutant embryos for both 10 kb binned and 50 kb binned data across replicate 1, replicate 2, and the combined replicates. **b**, Insulation profiles are shown for each biological replicate (replicate 1, orange line; replicate 2, blue line) for 50 kb and 10 kb binned data in wild-type embryos and DC mutant embryos. Insulation profiles are calculated using a 500 kb × 500 kb insulation square (10 bins × 10 bins for the 50 kb binned Hi-C data, and 50 bins × 50 bins for the 10 kb binned Hi-C data). The insulation profiles are consistent across replicates. Green tick marks, TAD boundaries identified using combined replicate data. **c**, Differential insulation plots derived from the insulation profiles calculated above (50 kb binned and 10 kb binned Hi-C data). **d**, 50 kb binned heatmap depicting the difference in chromatin interactions expressed as

the difference in Z-scores between wild-type and DC mutant. **e**, Plot showing the compartment analysis calculated using the 50 kb binned wild-type Hi-C data. A/B compartment profile was determined by principle component analysis. First Eigen Vector value representing compartments (black) is plotted along the chromosome, revealing three zones for each autosome: two outer sections and the middle third of the chromosome. Positive Eigen1 signals represent the B (inactive compartment) and negative Eigen1 signals represent the A (active compartment). The compartments at chromosome ends display increased interactions with each other, both in *cis* and in *trans* (see Extended Data Fig. 1a). Also shown is the average binding of the lamin-associated protein LEM-2 along the chromosomes (grey). Overall compartmentalization correlates with LEM-2 binding, showing that compartments at both ends of chromosome I are located near the nuclear periphery.
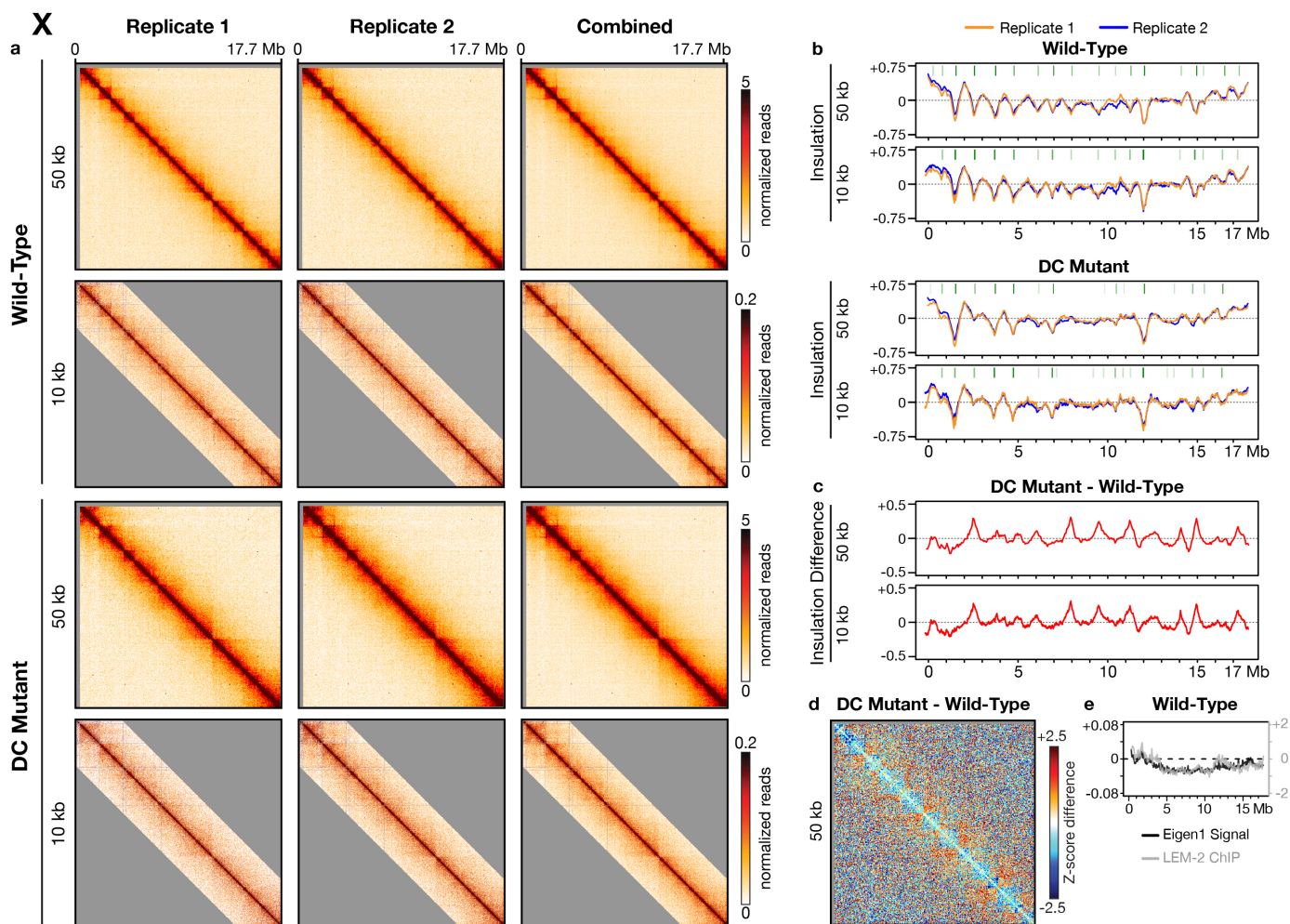
Extended Data Figure 5 | Compartment and insulation analysis for chromosome X in wild-type embryos and DC mutant embryos. a–e, See legend to Extended Data Fig. 4. In **e**, only two compartments are observed for chromosome X, compared to three for chromosome I. Overall compartmentalization correlates with LEM-2 binding, showing that the compartment at the left end of chromosome X is located near the nuclear periphery.
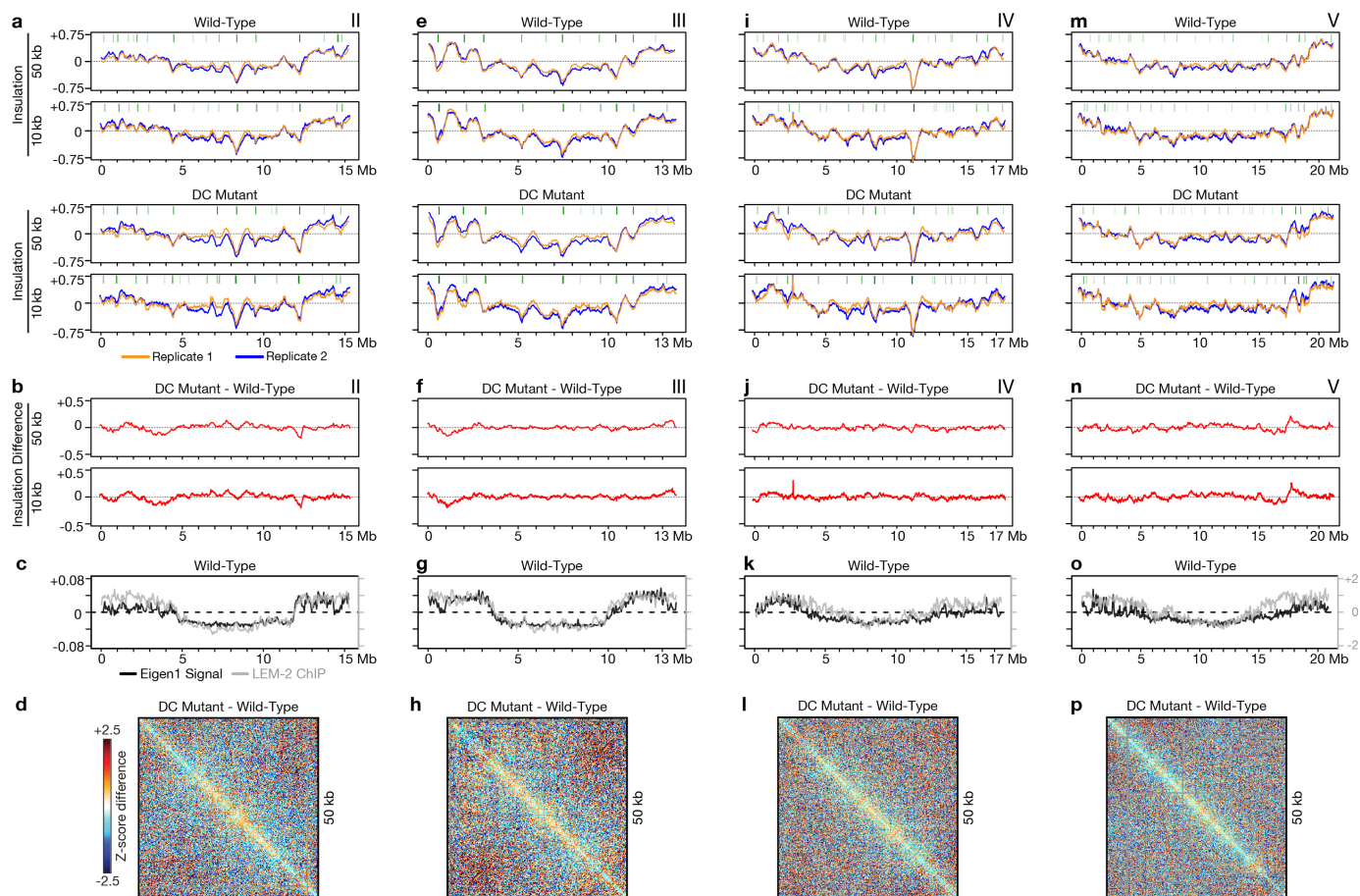
Extended Data Figure 6 | Compartment and insulation analysis for chromosomes II, III, IV and V in wild-type embryos and DC mutant embryos. a–d, Chromosome II. **e–h**, Chromosome III. **i–l**, Chromosome IV. **m–p**, Chromosome V. **a, e, i, m**, Insulation profiles for each biological replicate (replicate 1, orange line; replicate 2, blue line) for 50 kb or 10 kb binned Hi-C data in wild-type embryos and DC mutant embryos. Green lines, TAD boundaries identified from combined replicate data. **b, f, j, n**, Differential insulation plots made from insulation profiles (50 kb binned or 10 kb binned

Hi-C data). **c, g, k, o**, Plots show chromosome compartment analysis calculated with 50 kb binned data. Average binding of the lamin-associated protein LEM-2 is shown along the chromosomes (grey). Compartmentalization correlates with LEM-2 binding; compartments at both ends of autosomes are near the nuclear periphery. **d, h, l, p**, Heatmaps 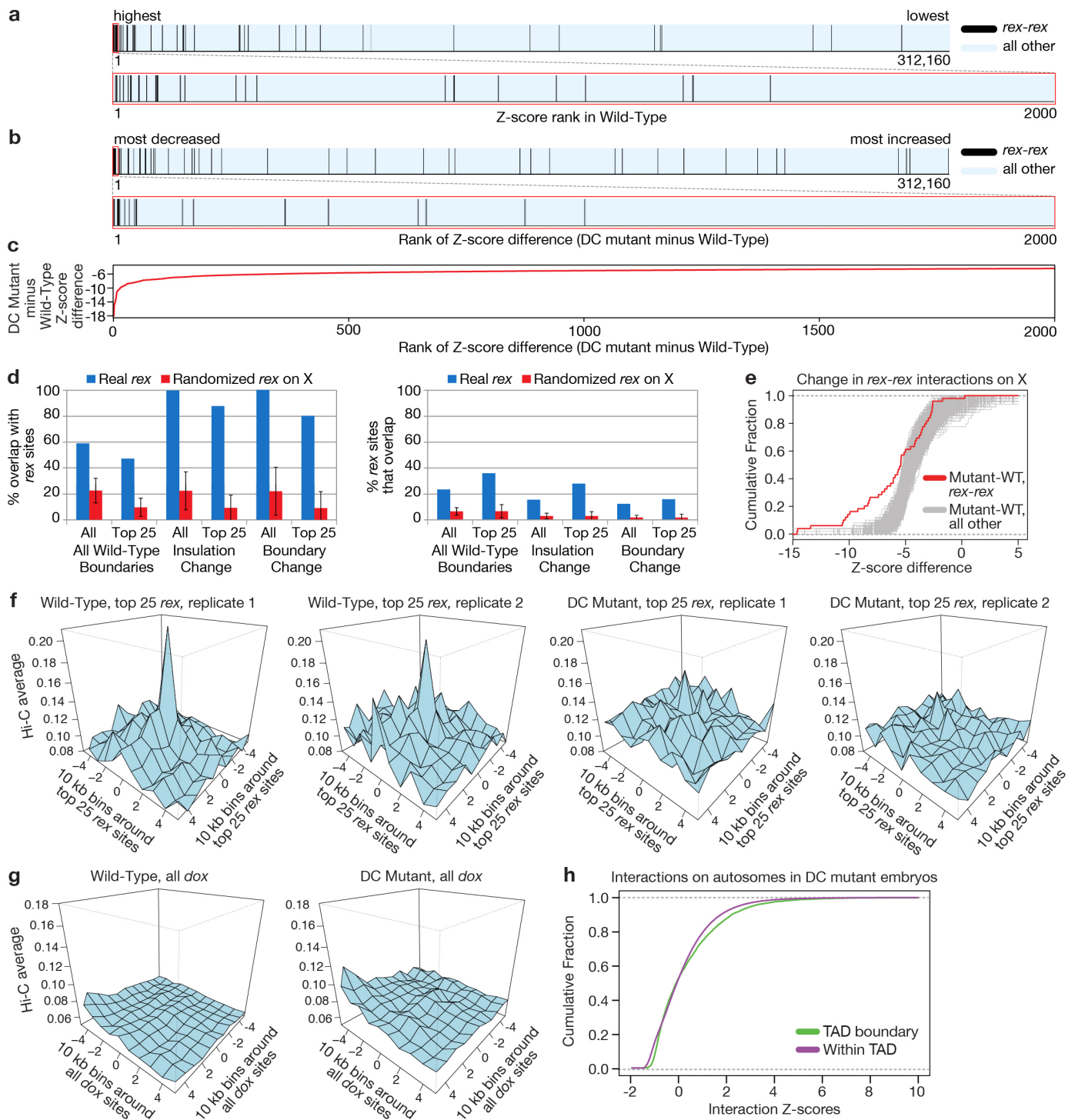(50 kb bins) show differences in chromatin interactions as the differences in Z-scores (DC mutant minus wild-type embryos).

Extended Data Figure 7 | *rex* sites are enriched at TAD boundaries and in top Hi-C interactions. a, Tick plots rank the interaction Z-scores for the top 25 highest-affinity *rex* sites (black) among all other 10 kb bin Hi-C interactions on chromosome X (light blue). Bottom plot amplifies top 2,000 interactions. Density of black ticks (left) shows strong enrichment of *rex*–*rex* interactions among the most significant chromosome X interactions. **b**, Tick plots rank the Z-score differences (DC mutant minus wild-type embryos) for interactions between the top 25 *rex* sites among all other differences on chromosome X. Bottom plot amplifies top 2,000 changes. **c**, Quantification of Z-score differences for top 2,000 changes in (**b**). **d**, Bar graphs depict overlap between chromosome X TAD boundaries and *rex* sites. Three sets of TAD boundaries are shown: all 17 boundaries; 8 boundaries with an insulation change (DC mutant minus wild-type) >0.1; 5 boundaries present in wild-type embryos but absent in DC mutants. Overlap is calculated for the entire set of *rex* sites or just the top 25 *rex* sites. Percent of boundaries that overlap *rex* sites (left). Percent of *rex* sites that overlap each set

of boundaries (right). Red bars, same sets of overlaps were calculated with 1,000 random sets of *rex* site positions along chromosome X. Average overlap and standard deviation are shown. No randomized set had as much overlap as the true *rex* set (*P* < 0.001). **e**, Cumulative comparison of Z-score differences for *rex* interactions and for 1,000 randomized sets of non-*rex* interactions (same number as in *rex* set). These *rex* or non-*rex* interactions had Z-scores >4 in wild-type embryos. *rex* interactions are reduced more in DC mutants than other similarly strong chromosome X interactions (*P* = 0.037; *rex*-interaction differences were significantly more reduced (KS test) than random interaction sets for 963 of 1,000 cases). **f**, 3D plots of Hi-C interaction profiles (normalized read counts) around top 25 *rex* sites for 2 Hi-C replicates of wild-type embryos or DC mutants. **g**, 3D plots of interactions between *dox* sites in wild-type embryos and DC mutants show no interaction peak. **h**, Cumulative plots show no difference in DC mutants for the distribution of autosomal Hi-C interaction Z-scores (10 kb bins) in TADs or at boundaries.
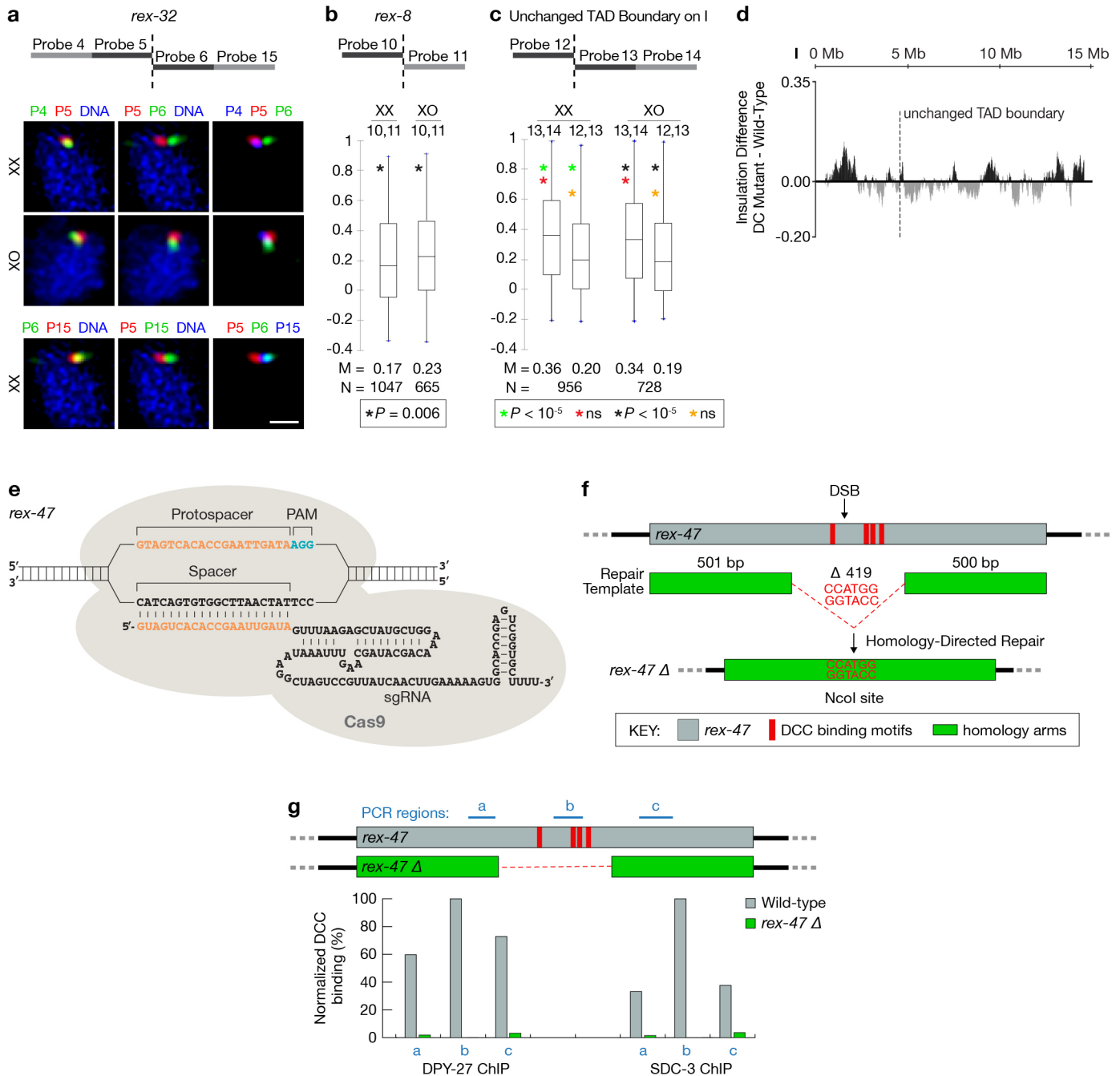
Extended Data Figure 8 | Visualization and disruption of TAD boundaries.
a–d, Visualization of DCC-dependent TAD boundaries in single cells confirms Hi-C analysis. **a**, Representative confocal images of embryonic nuclei of different genotypes stained with a DNA intercalating dye (blue) and FISH probes surrounding *rex-32*. Scale bar, 1 μm. **b**, Quantification of colocalization between FISH probes flanking *rex-8* (see Fig. 2a) in XX and XO embryos confirms the DCC-dependent boundary identified by Hi-C. Because TADs on either side of *rex-8* are small, we could only use one 500 kb FISH probe for each TAD. **c**, Quantification of colocalization between FISH probes for a TAD boundary on chromosome I (dashed line in **d**) in XX and XO embryos confirms the DCC-independent boundary identified by Hi-C. **b**, **c**, Box plots show the distribution of Pearson's correlation coefficients between pairwise combinations of FISH probes. Boxes represent the middle 50% of coefficients, and the central bar within indicates the median coefficients (M). *N*, total number of nuclei. *P* values derived using the one-tailed Mann–Whitney *U*-test are shown below each graph. NS, not significant. **d**, Insulation difference plot of chromosome I for DC mutant insulation profile minus wild-type insulation profile. **e–g**, Deletion of endogenous *rex-47* by Cas9 disrupts DCC binding and TAD boundary formation. **e**, Schematic illustration of the sgRNA–Cas9 complex interacting with the *rex-47* target sequence. **f**, Cas9-mediated deletion of *rex-47*. Top, diagram showing the location of DCC binding motifs within *rex-47* (red bars) and Cas9-induced double strand break (arrow). Middle, diagram of the double-stranded repair template containing two ~500 bp homology arms and an NcoI restriction site. Bottom, after precise homology-directed repair, a 419 bp region containing all DCC binding motifs was deleted and replaced with NcoI. **g**, Loss of DCC binding at endogenous locus carrying the *rex-47* deletion. DCC binding at three ~100 bp regions located upstream (a), within (b) or downstream (c) of the 419 bp deletion was examined using ChIP–qPCR. Histogram shows the ChIP–qPCR signal for DCC components DPY-27 or SDC-3 at target regions relative to the level at region b in wild-type embryos.
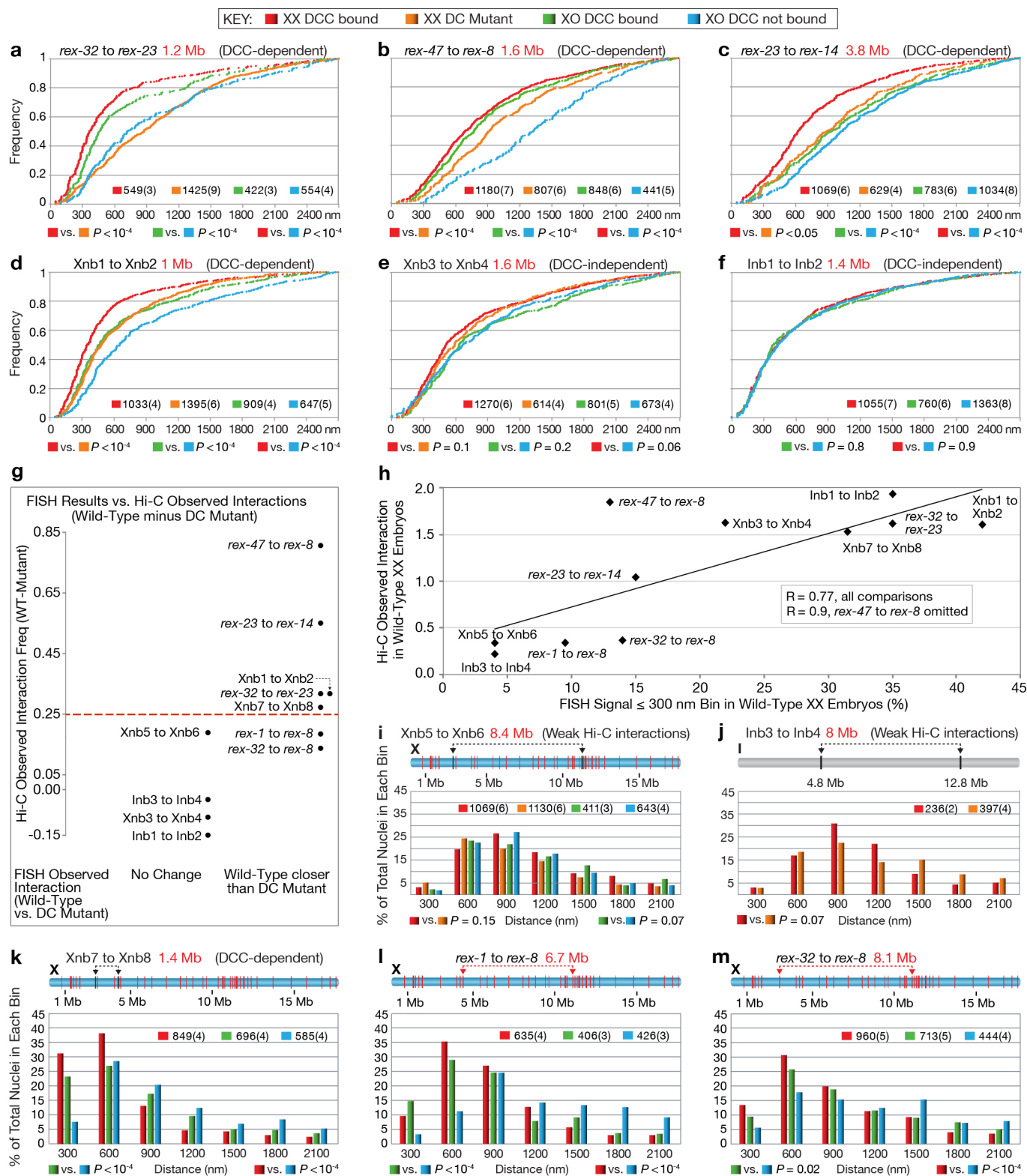
Extended Data Figure 9 | Quantitative FISH shows that *rex* sites colocalize more frequently if the DCC is bound to chromosome X. a–f, Data from histograms in Fig. 4b–g shown as cumulative plots. Number of nuclei and embryos (parentheses) assayed are shown (also for **i–m**). Distance between loci (red) and DCC dependence or independence of Hi-C interactions (black) are shown. *P* values (chi-squared test) compare values in the 0–300 nm bin to those in 301–2,700 nm bins. Same statistical analysis for (**i–m**). **g**, Correlation between DCC-dependent Hi-C interactions and DCC-dependent FISH colocalization. *y* axis, difference between wild-type and DC mutant Hi-C observed interaction frequency at 50 kb resolution. Higher number shows greater DCC-dependence. *x* axis shows two categories defined by FISH: sites with unchanged colocalization frequency in DC mutant (DCC-independent) (left); sites with less frequent colocalization in a DC mutant (DCC-dependent) (right). Red dotted line, cutoff for calling a Hi-C interaction 'changed' between the wild type and DC mutant. **h**, Scatter plot shows correlation between Hi-C and FISH data. *y* axis, Hi-C observed interaction frequency in 50 kb bins. *x* axis, percentage colocalization (that is, 300 nm bin) by FISH. $R = 0.77$ for all comparisons; $R = 0.9$ if the *rex-47*–*rex-8* interaction is omitted. **i–m**, Histograms show quantification of 3D distances between two FISH probes. **i, j**, Distant loci on chromosome X or chromosome I with weak Hi-C interactions. **k**, DCC-dependent interaction between X sites lacking DCC binding. **l–m**, DCC-dependent interactions between distant *rex* sites.

a Wild-Type Gene Expression on X, RNA-seq
b DC Mutant Gene Expression on X, RNA-seq
c Gene Expression Change on X, RNA-seq

d Wild-Type Gene Expression on X, GRO-seq
e DC Mutant Gene Expression on X, GRO-seq
f Gene Expression Change on X, GRO-seq

a-f: *Significance did not withstand multiple testing correction.*

g Expression changes for genes within TADs, RNA-seq

h Expression changes for genes within TADs, GRO-seq

i Changed boundaries on X, RNA-seq — Changed boundaries on X, GRO-seq

Distance from center of changed boundaries (10 kb bins)

Extended Data Figure 10 | DCC-dependent TADs influence global rather than local gene expression. Gene expression analysis was assayed using RNA-seq or GRO-seq, as indicated. **a**, **b**, Boxplots depict expression levels for wild-type or DC mutant embryos assayed by RNA-seq for chromosome X genes at changed TAD boundaries, unchanged TAD boundaries, all TAD boundaries or genes not at TAD boundaries. Expression levels are given as normalized read number per kilobase of gene length. **c**, Boxplots depict the fold change in expression assayed by RNA-seq between wild-type embryos and DC mutant embryos for genes at changed TAD boundaries, unchanged TAD boundaries, all TAD boundaries or genes not at boundaries. The lowest-expressing genes (bottom 10%) were removed from analysis. **d**–**f**, As in **a**–**c**, but assayed by GRO-seq with gene expression levels given as fragments per kilobase of transcript per million mapped reads (FPKM). For **a**–**f**, *P* values were calculated using the Mann–Whitney *U*-test; significance did not withstand multiple testing correction. **g**, **h**, Boxplots depict the fold change in the gene expression between wild-type and DC mutant embryos based on RNA-seq or GRO-seq for chromsome X and chromosome I. Each box has genes from one TAD on chromosome X (left) or chromosome I (right). Lowest-expressing genes (bottom 10%) were removed from analysis. No discernible pattern was evident for expression changes versus gene location. **i**, Boxplots depict the fold change in chromosome X gene expression between wild-type embryos and DC mutant embryos relative to the distance from the TAD boundary. Each box contains genes in 10 kb bins radiating out from the centre of each TAD boundary. The lowest-expressing genes (bottom 10%) were removed from analysis. No discernible pattern to the gene expression changes exists, as assayed by RNA-seq (left) or GRO-seq (right). Weak significance and lack of concordance between RNA-seq and GRO-seq data suggest no biologically relevant correlation between TAD boundaries and local regulation of gene expression.
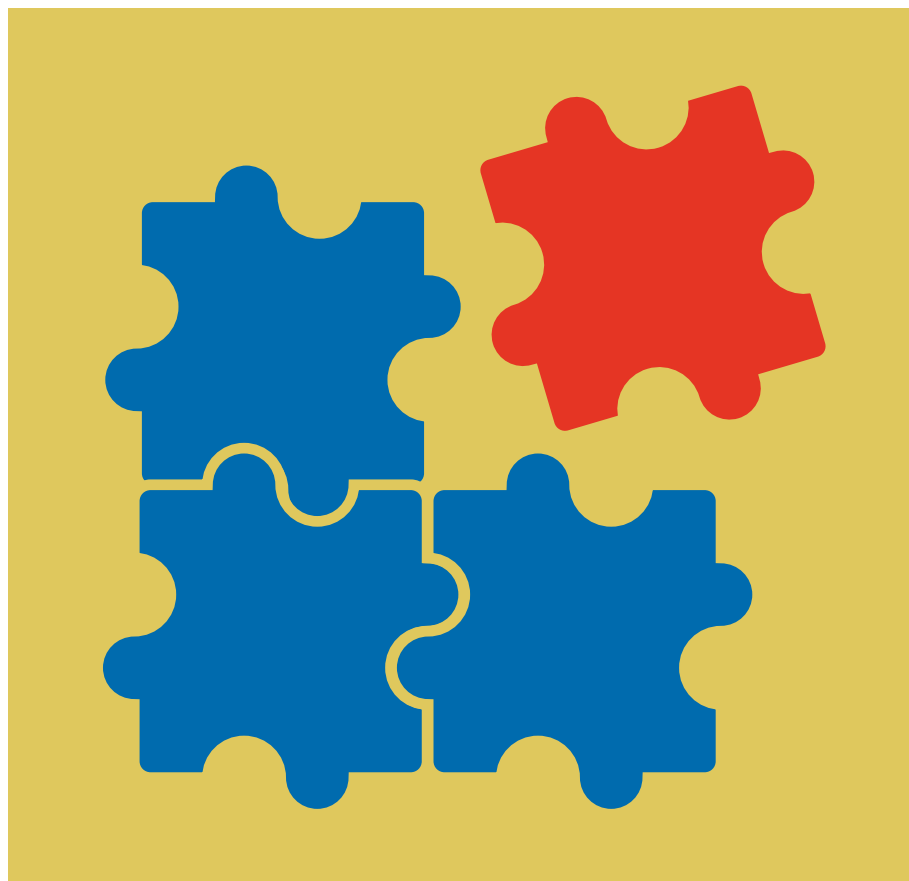
CAREERS

NATUREJOBS For the latest career listings and advice www.naturejobs.com

COLLABORATIONS

Recipe for a team

A scientific collaboration is vulnerable to derailment unless members learn to trust each other at the outset.

BY VIRGINIA GEWIN

Marine biologist Benjamin Halpern was part of an 11-person team that met up in 2012 at an eco-resort on the southern tip of Australia's Great Barrier Reef. The group's mission was to develop a scientific method able to identify species-conservation solutions that could minimize costs without unduly affecting any specific group of people. Each morning for a week, the team debated and discussed data, modelling and statistics.

In the afternoons, they all went snorkelling, scuba-diving or birdwatching together. Team members had brought spouses, partners and children, and the meeting felt as much like a social seashore soirée as it did a scientific collaboration. "We got to see many different sides of our colleagues, which I think helped everyone bond more," says Halpern.

By working and playing hard together for a full week at the project's outset, group members built the connections and trust that were needed to share their ideas and develop new ones together. Within weeks, the team had submitted its findings on effective conservation planning,

and these were published just three months later (B. S. Halpern *et al. Proc. Natl Acad. Sci. USA* **110,** 6229–6234; 2013). Since then, various members of the group have secured further funding to expand their work and to bring in new collaborators, says Halpern, of the Bren School of Environmental Science and Management at the University of California, Santa Barbara (UCSB). He has participated in about 20 collaborative efforts supported by the National Center for Ecological Analysis and Synthesis at UCSB, an ecology think tank that funds team-oriented interdisciplinary projects. "Good ideas are relatively cheap; it's the execution of them that is hard," says Halpern. "What makes a collaboration succeed or fail is having the right team."

Not every collaborative posse can meet, as Halpern's did, in a luxurious location to forge ties. But group members can take steps to get their project off on the right foot — and to keep it moving forward. They need to take those steps because funding schemes increasingly encourage or even require collaboration, says Koen Frenken, who teaches innovation studies at the University of Utrecht in the Netherlands. It is especially important for junior researchers to do all that they can to ensure that their group — and their standing in it — remains on track.

Word about who the 'good' collaborators are spreads quickly. These people are highly sought after, whereas 'bad' collaborators may never learn about their own unfavourable reputation (see 'Caricatures'). "Academic communities are quite small and people want to avoid conflict," says Linus Dahlander, who studies collaborations at the European School of Management and Technology in Berlin. Most researchers who become frustrated with an ineffective team member never talk about it — they simply do the slacker colleague's work, give them credit and then avoid partnering with them again, says Barry Bozeman, director of the Center for Organization Research and Design at Arizona State University in Phoenix.

FALLOUT WARNINGS

Despite everyone's best efforts, collaborations can fall apart for any number of reasons — misunderstandings, faulty assumptions or personality clashes. One team member could have a strong personality that dominates the others. More often, members assume that colleagues share their views. "Don't assume everyone knows what you know or perceives things the way you do," Bozeman says.

This is a particular problem with international collaborations, when cultural or language barriers can challenge a team. But there are also structural differences in such partnerships, says Melissa Anderson, a higher-education specialist at the University of Minnesota in Minneapolis, who researches the scientific-integrity aspects of teamwork. She says that those differences can include how collaborations are organized and financed in different nations, as well as the federal and national laws that govern the work of each team member in a different country. "Not all countries have exactly the same expectations regarding integrity issues," she says.

And there can be confusion about what constitutes plagiarism, or cultural differences that make it unclear how to address wrongdoing or how to challenge a superior, she adds. Team members can avoid many of these potential problems by making time to meet with one another to discuss the partnership's financial, ethical and cultural issues in person, she says.

Even collaborators from the same country can be derailed by an absence of face time, especially if they span different disciplines. The problem is worsening in the digital era, when scientists need not ever meet in the flesh to join up on a research project.

Steve Fiore experienced first-hand how important it is to make sure that common terms mean the same thing to everyone. He was part of a multi-discipline, multi-university effort in 2010 to develop teams of humans and robots, an endeavour that almost fell apart because of a simple word that had different meanings for everyone involved.

"We were spinning our wheels" on the development and testing of models, says Fiore, a cognitive scientist who studies group research at the University of Central Florida in Orlando. "Then I realized that 'model' meant different things to the engineers, to the computer scientists who were developing artificial intelligence and to the social scientists."

The confusion was exacerbated by the use of teleconferences for group meetings. "There was a lack of in-person cues that could have made the misunderstandings more apparent," he says. Once he realized what was happening, he explained the problem to the team and the collaboration regained momentum. Group members reviewed their discussions thereafter through e-mail to avert any repeat disasters.

SCIENTIFIC PRENUP

How can a collaboration be stopped from going sour? One way is to create the scientific equivalent of a prenuptial agreement (see 'Tricks for tackling teamwork'). In addition to defining team-member expectations, a 'prenup' spells out the overall goals and vision for the team and what constitutes authorship as well as communication and contingency plans.

Junior investigators might struggle to persuade more-senior collaborators to adopt this formalized approach, says Kara Hall, director of the Science of Team Science Team at the US National Institutes of Health (NIH) in Bethesda, Maryland. But they can at least initiate a conversation about the issues that are covered by such a document — determining authorship, for instance.

There are no data on whether the use of a research prenup is on the rise, but Hall says that requests for presentations that discuss the topic have skyrocketed. Collaboration veteran Halpern encourages each group, at minimum, to spend time talking about expectations and authorship and to consider writing down verbal agreements at the outset of every team project. One team with which he worked agreed that individual researchers who were passionate about publishing on a specific idea that had evolved from a group effort should be free to do so, without the expectation that every team member would be an author. As a result, more good work came from the team effort.

It can also help to draw a shared diagram that represents the research problem and every member's place in it, says Paul Hirsch, who studies interdisciplinary collaborative processes at the State University of New York (SUNY) in Syracuse. At the most basic level, team leaders and collaborators should discuss expectations, working styles and how to execute their shared vision. Individual collaborators can devise informal practices and rules that work for them, including collaboration-management procedures.

NIH ombudsman Howard Gadlin, who studied successful collaborations while he was co-authoring the NIH report *Collaboration and Team Science: A Field Guide* (2010), found that team members in successful collaborations had a common vision for the work that they were doing and how their contributions fit into the larger mission (see go.nature.com/ghcwfs). Effective team leadership can also help to ward off conflict. "Group leaders are responsible for creating a culture where people share ideas that benefit the team; otherwise, we get in a situation like Gollum in the *Lord of the Rings*, with no one sharing their 'precious' ideas," Dahlander says.

> *Word about who the 'good' collaborators are spreads quickly. These people are highly sought after.*

Not surprisingly, the promise of a big pot of funding can inspire sharing between team members. Last year, Chris Nomura of SUNY's Syracuse campus was one of a number of 'green' chemists and physicists who were assembled by the Research Foundations of SUNY in a bid to unite expertise in disparate fields across the university's four sites.

The challenge was to get this subgroup to spend a small amount of seed money to pursue joint research priorities connected with green composite materials. But no one knew anyone else, and some people were confused about why they had been selected. Several members, burned by previous collaborations, were wary of sharing their ideas lest they be stolen.

Nomura says that the subgroup signed a non-disclosure agreement before the first meeting so that they would feel comfortable talking openly. They also took time to discuss negative past experiences and they made a pact defining behaviours that should be avoided — chiefly, that any ideas discussed in the group would not be used in individual grant proposals without the permission of the group.

FIND COMMON GROUND

At the time, Hirsch advised Nomura's team to find a way to coalesce around one goal: to identify a shared research aim. Members of the group became collectively excited about developing innovative energy-efficient ways to produce composite materials, a scheme that, as it turned out, could successfully compete for funding from the New York State Energy Research and Development Authority. "Something pretty amazing happened after we talked about our research and decided to

---

PRACTICAL TIPS
Tricks for tackling teamwork

The secret to a successful collaboration is forethought. The US National Cancer Institute's Team Science Toolkit offers a host of tips (go.nature.com/fyrefu). Below are some more thoughts gleaned from interviews for this article.
- Choose team members who are open to fresh ideas and willing to engage in a thoughtful manner.
- Team leaders should create an environment in which people can disagree constructively and in which there is freedom to ask 'stupid' questions.
- Any 'prenuptial' agreement on roles and responsibilities should be negotiated as a team at the start.
- Team leaders should assign products of the collaboration to the team members who will get the most career benefit from them.
- Junior researchers should organize teaching schedules to allow enough time for joint projects. **V.G.**

TEAMSHIRK
Caricatures

These are the stereotypes to avoid adopting in a collaboration if you wish to be welcomed into one again.

● **The overcommitted superstar.** The high-profile, highly sought-after researcher who lends wattage to the effort but who cannot offer much time or attention to an individual team.

● **The social loafer.** The team member who is simply not engaged — perhaps owing to a lack of shared vision or a lack of goal alignment.

● **The know-it-all.** The collaborator who dominates the conversation and does not make space for all colleagues to be heard.

● **The lurker.** The team member who withholds her or his own insights while absorbing everyone else's. The lurker is driven by tough competition but often burns bridges. **V.G.**

apply for grants," says Nomura. "We agreed on an unequal disbursement of the seed money — some groups got less money and some got more, realizing that strategically it would benefit us all in the long run."

Halpern reminds early-career researchers that what they lack in collaboration experience, they can make up for with time and energy. "Offering to contribute is the best way to get involved in collaborations — and possibly shift to the next phase of their career," he says. As a first-year graduate student on his first collaborative team, he offered to lead a meta-analysis of existing data on the conservation value of marine reserves. It was a transformative move that positioned him to work with a network of scientific leaders in marine conservation.

But despite the best efforts to maintain momentum, sometimes a collaboration simply has to be abandoned. A team can grow stale, like any relationship, or the obstacles can become too overwhelming. "I've seen collaborations that fell apart and never recovered," says Gadlin.

Ultimately, however, it is not success — as measured by the number of citations — that has the most substantial impact on the continuation of a collaboration. Often, the longevity of a team project can be judged by the beer test. "If collaborators don't like each other enough to go for a beer after the meeting, it can be a sign of pending doom," Dahlander says. ∎

Virginia Gewin is a freelance writer in Portland, Oregon.

COLUMN
Match that PhD

Lab leaders discuss how to find the perfect graduate student for a research group.

BY DEBORAH J. MARSH, KIRSTY FOSTER & CAROLYN D. SCOTT

Graduate students can consult reams of material on how to choose a PhD supervisor and select the best and most appropriate research group. But almost no resources exist for principal investigators (PIs) — especially those in the early stages of their own careers — on how to choose a PhD student for their lab or research team. How do these leaders decide who will be the best 'match'?

If you assume the role of supervisor, mentor or PI, you will provide much of the guidance and support that is crucial for a student's career development. Deciding whether to take on such a task requires much deliberation. You will need to consider whether your research group, project and academic environment will allow the student to flourish and receive the proper level of supervision, whether the student can develop the skills necessary to maximize your project's success and whether he or she will be a good fit with your group.

You will need to consult your team. Current members must feel confident that they share goals with their future colleague. As team leader, you will need to ensure that a new member will contribute to the group's work and will not adversely affect the team dynamic. Ask the applicant to talk to your team and find out what members think. You will probably learn about the applicant's research experience, communication and social skills and whether she or he prefers to work in a group or solo.

Setting an exercise for a PhD candidate can also prove useful for evaluating the student's research background and writing and problem-solving skills. We routinely ask candidates to choose and critique one of our published papers and to suggest how the study could be improved. The choice of paper provides clues about the student's interests, and we learn about his or her knowledge of the field, and ability to organize and communicate ideas. We have also found that the task both attracts and dissuades candidates. Once, after assigning it, we did not hear again from the candidate. Other candidates have dived in. "It showed that you cared what I thought," one student told us after completing it.

You should also ask applicants why they want a PhD, why they are interested in your group, which research discovery they are most proud of and what comes most easily to them, whether it be benchwork, fieldwork or something else. Applicants' answers provide information about their attitudes and aptitudes. For example, a student who expresses a preference for data analysis might be best suited to a project that involves extensive statistical or bioinformatic analyses.

Many PhD students want to be asked specific questions. Our students, for example, have indicated that they think that we should ask about evidence of positive relationships with previous supervisors or lecturers, a strong academic record, an ability to work well in a team environment and curiosity about and enthusiasm for their research areas.

Most students are highly motivated to succeed. Great achievement generally takes place in an environment of high standards, so you will need to discuss your expectations. These could include attending conferences, adhering to agreed milestones and participating in seminars and journal clubs.

Choosing the right PhD students for a team is more important than ever if we, as supervisors and mentors, are to make a positive impact on the scientific endeavours that will be led by those whom we train today. ∎

Deborah J. Marsh is a professor of molecular oncology, **Kirsty Foster** is an associate professor in medical education and **Carolyn D. Scott** is a sub dean in postgraduate training at the Kolling Institute of Medical Research, University of Sydney, Australia.

THE RAVELLED SLEEVE OF CARE

Ties that bind.

BY ANATOLY BELILOVSKY

Dark letters danced in the amber glow, and Vera rubbed her temples to push back the migraine that threatened to ambush her from the flickering edges of her vision. She closed her eyes for a moment, then opened them again. The error was obvious; in her mind she saw the microprocessor execute the instruction, overwriting program memory with data and creating an abnormal loop. Another moment of contemplation produced a workaround. Vera ran through the instruction sets in her mind again; the operations marched in neat, obedient rows. She had no doubt that the patch would work. A smile crept across her face as she began to write:
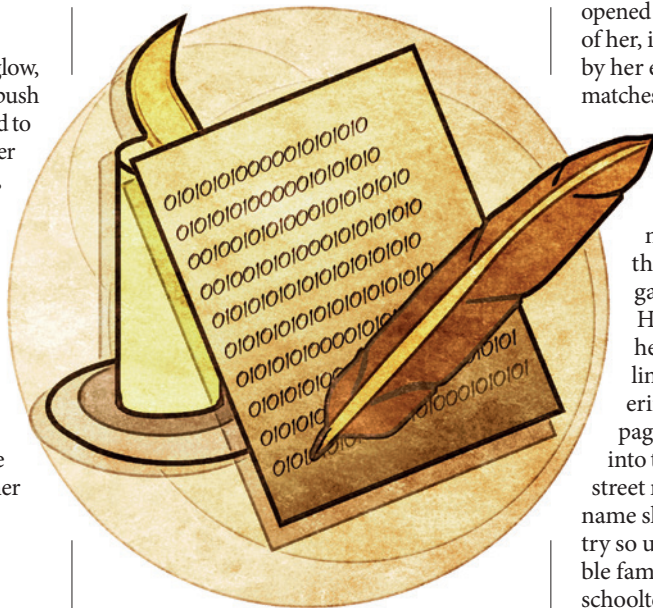
"Dear Cousin Grisha,
  Many thanks for the medicine you sent Grandma Liza through the Aeroflot crew; whatever bribes you gave them were worth it. Grandma's ankles are so much less swollen than they used to be, and she no longer faints every five minutes the way she used to on the old medication."

Vera heard a goose honk in the birdhouse next to the frozen pond, and she stood up to reach for her shotgun, but no other bird cried in answer. She sighed in relief and turned towards her daughter's crib. Little Anechka wriggled in her sleep, stretching her arms; her lips, barely visible in the darkness, pursed momentarily, then relaxed. Vera rocked the crib, gently, and bent back to her task.

"My sister Valentina, as you know, had her baby right after I had mine. She lost much blood while giving birth, and needs special diet to restore her health. Meat is very expensive, but goose liver and kidneys are still available (to foxes and stoats, much of the time, but still), and of course in the fall there were mushrooms so we are not starving. Do you remember how we went mushroom picking as children? Chanterelles and slippery jacks grew under every oak, we'd come back groaning under their weight after an hour or two — though the forest you remember from our childhood is off limits, now that we know what really happened at Kyshtym, so we have to take the suburban train and ride for two hours in the opposite direction till we get upriver from that awful place. Valentina's baby is very fussy, and I thank God every day that Anechka is growing well and sleeps through half the night now."

Vera's fingers grew stiff, and she rubbed her hands together for warmth, leaning to see if Anya's blanket still covered her. Close up, Anya's breath warmed her cheeks, tickled her eyelashes. She blinked and returned to writing.

"I am very happy to help you with your programming job. I read the machine language manual you sent me, and saw the memory dump and the screen image printouts, and I think I see a way to fix it. If you get the opportunity again, some vitamins with iron for Valentina would be most appreciated, and if American doctors know of anything better than milk of magnesia for Uncle Vanya's ulcers, I hope you can send some of that along as well. Here is my own small service to you, the patch you need to put in the program file, starting at Position 43217."

She closed her eyes again, marshalling rows and columns of symbols she saw in her mind's eye as clearly as if they were painted on her walls and ceiling. She saw the program execution tree as clearly in her mind as the view outside the window of her izba, as the streets (muddy till the recent frosts but now quite passable) of her village, as Anya's angelic face.

She took a deep breath, let it out, and opened her eyes. The candle guttered in front of her, its amber flame nearly extinguished by her exhalation, and she reached for the matches but the light returned, steadied, and brightened enough for Vera to see it belch a cloud of sooty, acrid smoke. She dipped her goose-quill pen into the inkwell that stood near the candle, touched the nib to the inkwell rim to drain excess oak-gall ink, and bent to the paper again. Hexadecimal characters marched from her mind, down her pen and onto the lined copybook paper, dancing in flickering amber light, and as she filled each page she folded it carefully and placed it into the envelope bearing an address on a street named after a shrub, in a city whose name she could not pronounce, in a country so unimaginably rich that even a humble family composed of an engineer and a schoolteacher could own a car, a home and, incredibly, a personal computer.

By the time she had finished, the Moon had sunk nearly to the horizon, adding through the window its silver glint to the amber-gold spark of the candle. She skipped two lines and wrote, in careful schoolgirl cursive:

"So please accept, with this letter, my most sincere wishes for happiness for your birthday, and continued success in your occupation, and for your children's perfect marks in their primary school, and for your wife to learn Russian so she can understand you better. I do not want to say goodbye, but it will soon be time for Anechka to wake, and I must bring water from the well and chop wood for our hearth, but please know that as always I remain,
  very truly,
    your loving cousin Vera."

She licked the envelope flap and pressed it shut. She counted out the stamps and licked them, too, before attaching them to the envelope and slipping it in her pocket. She checked Anechka's breathing one last time, then donned her greatcoat and picked up the water bucket to go outside. ■

Anatoly Belilovsky *was born in what is now Ukraine, learned English from* Star Trek *reruns, worked his way through a US college by teaching Russian while majoring in chemistry, and has, for the past 25 years, been a paediatrician in New York, in a practice where English is the fourth most commonly spoken language.*

ILLUSTRATION BY JACEY